AWARD NUMBER: W81XWH-16-1-0195

TITLE: Biological Characterization and Clinical Utilization of Metastatic Prostate Cancer-Associated lincRNA SchLAP1

PRINCIPAL INVESTIGATOR: Lanbo Xiao

CONTRACTING ORGANIZATION: Regents of the University of Michigan
Ann Arbor, Michgian 48109-1274

REPORT DATE: July 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2017	Annual	15 Jun 2016 - 14 Jun 2017

4. TITLE AND SUBTITLE

Biological Characterization and Clinical Utilization of Metastatic Prostate Cancer-Associated lincRNA SchLAP1

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0195

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Lanbo Xiao, Ph.D.

E-Mail: lanbox@med.umich.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Regents of the University of
Michigan
Ann Arbor, MI 48109-1340

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

During the research period, we followed our proposal to characterize the biological function of SChLAP1 in prostate cancer and assess the translational potential of SChLAP1 in prostate cancer. To ascertain the potential protein binding sites of SChLAP1, we performed PIP-seq in prostate cancer cells, and found SChLAP1 Exon 5 is the most protected exonic region. Then, we performed smFISH assay to precisely characterize the subcellular localization pattern of SChLAP1 at single molecule level. We found that SChLAP1 is highly nuclear specific transcript, and the absolute counts of SChLAP1 molecules was ascertained in prostate cancer cell line as well as prostate cancer biopsies. To functional annotate the interaction between SChLAP1 and SWI/SNF complex, we performed MNase-seq to determine the effect of SChLAP1 on global nucleosomes occupancy changes, we found that overexpression of SChLAP1 in prostate cells significantly affects nucleosome position around the genes with FPKM >1. Gene ontology (GO) analysis suggest that cell-cell adherens and cell cycle regulation are the most affected gene sets by SChLAP1. To sum up, our recently findings have substantial impact on understanding the biological function of SChLAP1 in prostate cancer progression, and helped us to further explore the potential of SChLAP1 to serve as biomarkers and therapeutic target for metastatic prostate cancer.

15. SUBJECT TERMS
LincRNA, SChLAP1, SWI/SNF, EZH2, ASO

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	47	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

Prostate cancer is one of the leading causes of cancer-related deaths in the United States among men and the most commonly diagnosed cancer in American males [6]. Most prostate cancer-related deaths are due to advanced disease. Identifying novel biomarkers for advanced prostate cancer and uncovering the molecular mechanism of disease progression will significantly benefit prostate cancer patients. Our previous studies defined *SChLAP1* (Second Chromosome Locus Associated with Prostate-1, *SChLAP1*) as a prognostic biomarker of advanced prostate cancer, and also showed that *SChLAP1* promotes prostate cancer migration and invasion by interacting with SWI/SNF complex. We hypothesize that elucidating ***the mechanism of SChLAP1-mediated abrogation of SWI/SNF function in prostate cancer progression will extend our knowledge of prostate cancer biology, and more importantly, reveal novel therapeutic targets against SChLAP1 and/or its downstream factors. To test our hypothesis, we propose the following specific aims:***

1) **To investigate the precise interaction between *SChLAP1* and SWI/SNF complex components.**
2) **To study the effects of *SChLAP1* on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer.**
3) **To determine the potential clinical utility of targeting *SChLAP1*, *SChLAP1*-SWI/SNF interaction and/or downstream effectors as a treatment modality in metastatic prostate cancer.**

1. KEY WORDS

LincRNA, SChLAP1, SWI/SNF, EZH2, ASO

2. ACCOMPLISHMENTS

What were the major goals of the project?

Following three aims were proposed.

Specific Aim-1: To investigate the precise interaction between *SChLAP1* and SWI/SNF complex components.

Specific Aim-2: To examine the effects of *SChLAP1* on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer.

Specific Aim-3: To determine the potential clinical utility of targeting *SChLAP1*, *SChLAP1*-SWI/SNF interaction and downstream histone modification as a treatment modality in metastatic prostate cancer.

What was accomplished under these goals?

Various tasks proposed in individual specific aims are written below. Results obtained under each task is shown in form of figures.

Specific Aim-1: To investigate the precise interaction between *SChLAP1* and SWI/SNF complex components.

Task 1: To Profile the potential protein binding (protecting) sites of *SChLAP1* by PIP-seq

We performed the PIP-seq in LNCaP cells to unbiasedly characterize the protein binding sites of SChLAP1 transcripts. As shown in **Figure 1**, crossing entire SChLAP1 transcript, we profiled multiple protected sites based on PIP-seq. Among those protested sites, we found Exon 5 of *SChLAP1* is one of the most highly protected exonic region, which suggested that Exon 5 of SChLAP1 might be the critical protein binding region. Interestingly, we also found one highly protected intronic region, which suggest the nascent SChLAP1 RNA transcript might also be bound by protein.

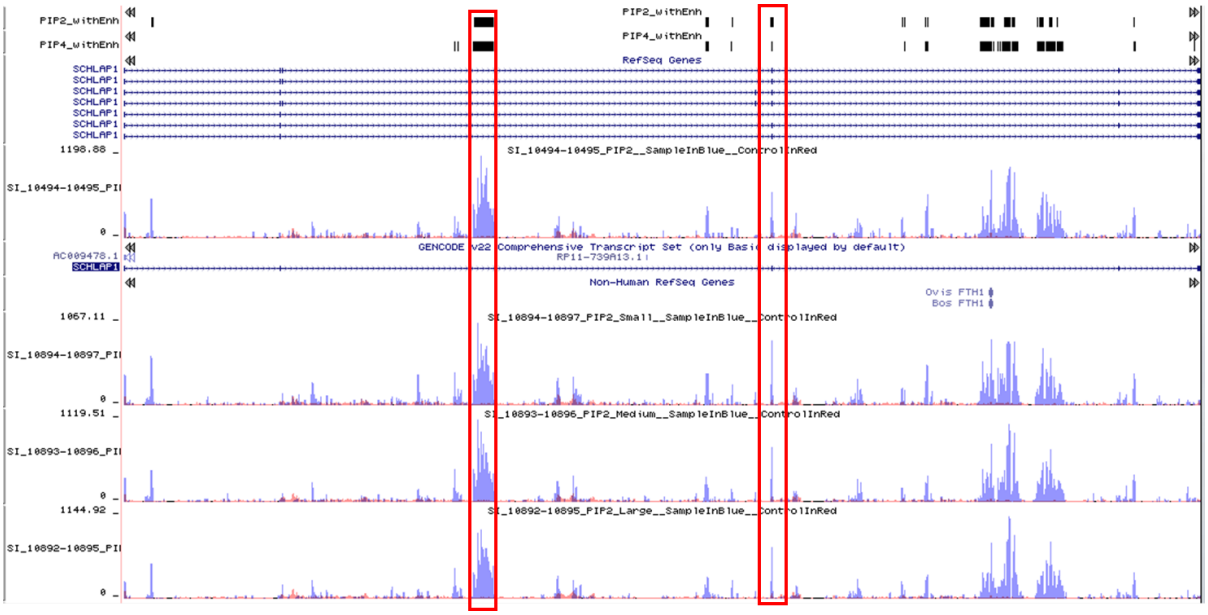


Figure 1. Genome browser view of PIP-seq result of SChLAP1 transcripts. Intronic and exonic protected regions were highlighted by red.

Task 2: To validate the potential protein binding (protecting) sites of SChLAP1 in multiple prostate cancer cells based on PIP-seq

We employed qPCR in LNCaP and VCaP cells to validate the potential intronic and protein binding (protecting) sites of *SChLAP1*. RNase I specifically digests single strand RNA, and RNase

III specifically digests double-strand RNA degradation. As shown in **Figure 2A**, for Exon 5 region, formaldehyde-fixed LNCaP and VCaP RNA could only be digested by RNase I but not RNase III, suggesting protected Exon 5 is not due to double-strand RNA formation, but might due to protein-RNA interaction. Similar to the Exon 5 region, the protected intronic region of SChLAP1 also was confirmed by qPCR (**Figure 2B**).

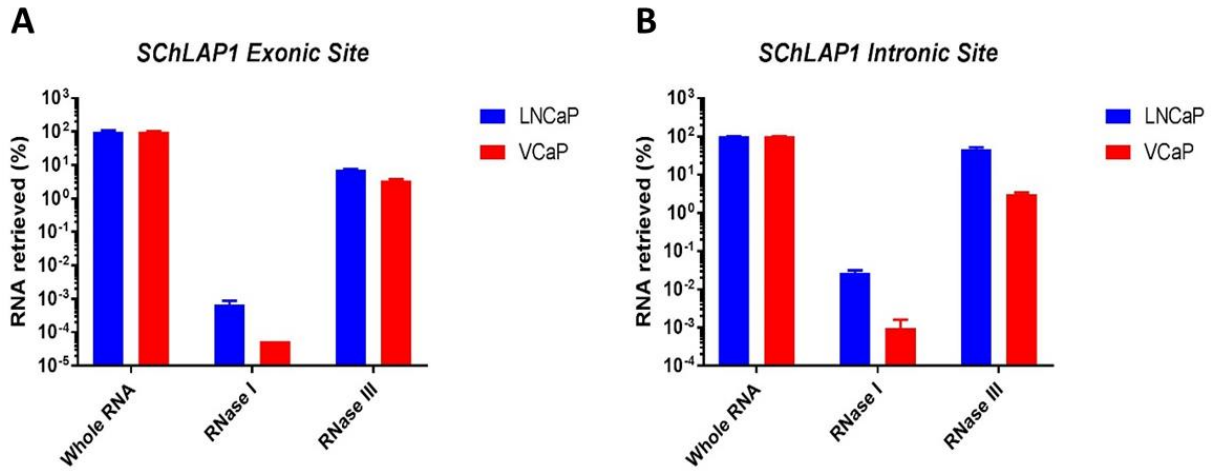


Figure 2. RNA of LNCaP and VCaP was treated by RNaseI and Rnase III. Levels of SChLAP1 exonic and intronic protected sites was quantified by qRT-PCR.

Task 3: To precisely characterize the subcellular localization of SChLAP1 in prostate cancer cells

To precisely characterize the subcellular localization of SChLAP1 in prostate cancer, we performed single molecular Fluorescence in situ hybridization (smFISH) for SChLAP1 in LNCaP cells. Comparing to conventional Fluorescence in situ hybridization (FISH), the smFISH assay is an absolute quantitative manner for RNA molecules at singe cell level. As shown in **Figure 3A**, we harnessed SChLAP1 exon specific smFISH probes (Red) and intron specific smFISH probes (Green) to SChLAP1 smFISH assay, and found SChLAP1 is highly nuclear specific transcript. By quantifying the absolute counts of each SChLAP1 molecules, we ascertained that absolute expression of SChLAP1 in LNCaP cells (**Figure 3B**).

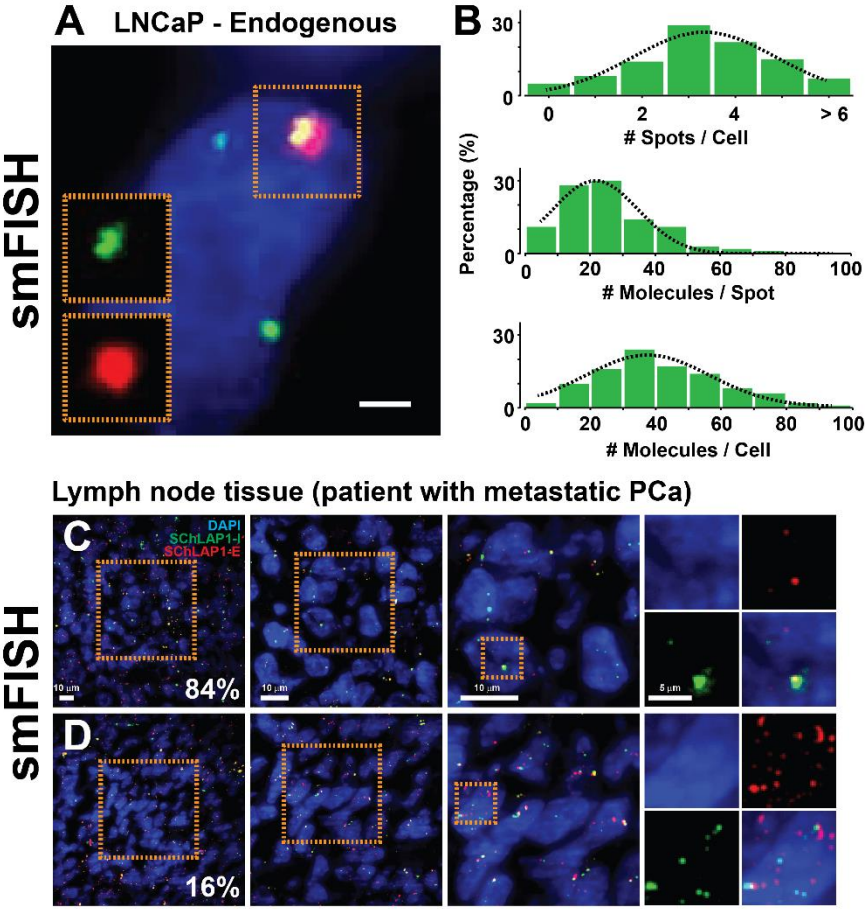


Figure 3 Localization of SChLAP1 by smFISH. A, smFISH for endogenous SChLAP1 (E-exons, I-introns) in LNCap Cell. B, Quantification of SChLAP1 expression. C and D, smFISH for endogenous SChLAP1 (E-exons, I-introns) in tissue (metastatic site). Percentage depicts fraction of samples with <6 spots (C) or > 6 spots (D) per cells.

Task 4: To validate the subcellular localization of SChLAP1 in prostate cancer patient's biopsies

we further employed smFISH approach to access the subcellular localization of SChLAP1 in metastatic prostate cancer biopsies. As shown in **Figure 3C**, we found the subcellular localization pattern of SChLAP1 in prostate cancer biopsies is consistent in both prostate cancer cell lines.

Task 5: To determine the co-localization of SChLAP1 and SWI/SNF complex in situ by immunofluorescence-FISH.

After we determined the subcellular localization of SChLAP1 in prostate cancer cell line and tissue, we assessed the co-localization of SChLAP1 with BRG or BRM in LNCaP cells. As shown in **Figure 4**, we successfully performed the immunofluorescence-FISH for SChLAP1 (Red) and BRG/BRM(Green). Since the BRG and BRM are abundant in LNCaP cells, it's difficult to claim that SChLAP1 is co-localized with SWI/SNF complex.

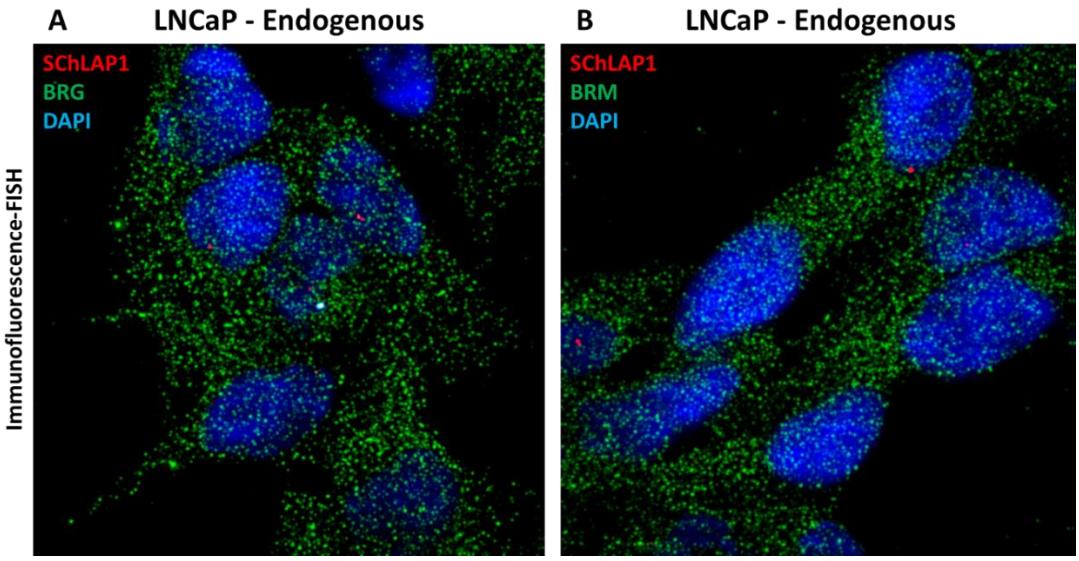


Figure 4: Co-localization of SChLAP1 and BRG or BRM complex in LNCaP cells by immunofluorescence-FISH. A, Co-localization of SChLAP1 and BRG in LNCaP cells. B, Co-localization of SChLAP1 and BRM in LNCaP cells.

Specific Aim-2: To examine the effects of *SChLAP1* on genome-wide nucleosome occupancy and H3K27Me3 in metastatic prostate cancer.

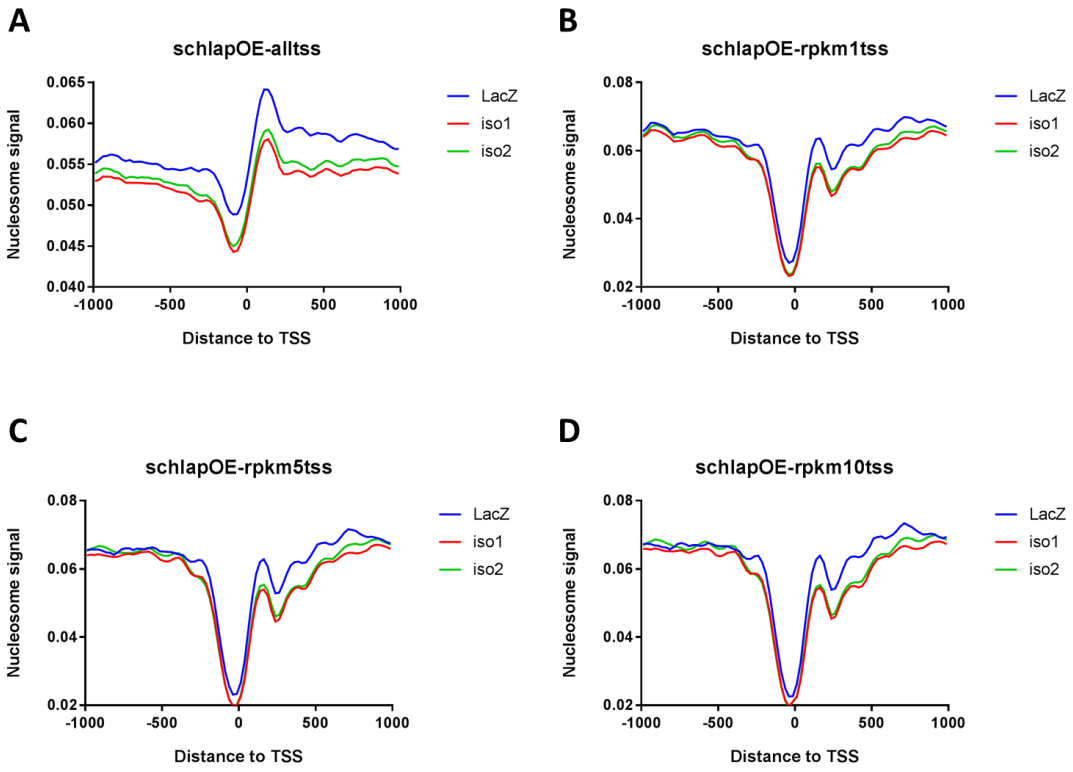


Figure 5. TSS-aligned overlay of nucleosome occupancy of RWPE-LacZ, RWPE-SChLAP1 iso1, and RWPE-SChLAP1 iso2 cells. Genome-wide nucleosome occupancy was determined by MNase-seq. A, TSS-aligned overlay of nucleosome occupancy of all TSS. B, TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=1). C, TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=5). D, TSS-aligned overlay of nucleosome occupancy of TSS (FPKM >=10).

Task 1: To profile the SChLAP1 overexpression-mediated nucleosome occupancy changes by MNase-seq analysis

To profile the SChLAP1-mediated nucleosome positions changes of surrounding transcriptional start sites (TSS), we stably overexpressed two SChLAP1 isoform in RWPE cells, and performed MNase-seq in the SChLAP1 overexpressing cells and parental cells. As shown in **Figure 5A**, the nucleosome positioning of all TSS were not significantly changed by SChLAP1. But for the TSS of highly expressed transcripts contained a nucleosome positioning changes at downstream of

TSS (**Figure 5B, 5C and 5D**).

Task 2: Integrative analysis of SChLAP1 overexpression-mediated nucleosome and transcriptomic changes.

To further functionally annotate the SChLAP1-mediated nucleosome position changes, all the genes were divided into six different groups based their SChLAP1-mediated TSS nucleosome positioning patterns (**Figure 6A and 6B**). For genes belong to group 1, 2 and 3, the TSS contained a nucleosome-depleted region (NDR) at downstream; for genes belong to group 4, 5 and 6, the TSS contained nucleosome-gain region (NGR) at downstream. Functional annotation analysis was performed for genes containing NDR or NGR, and the most significant relevant signature was showed (**Figure 6C**).

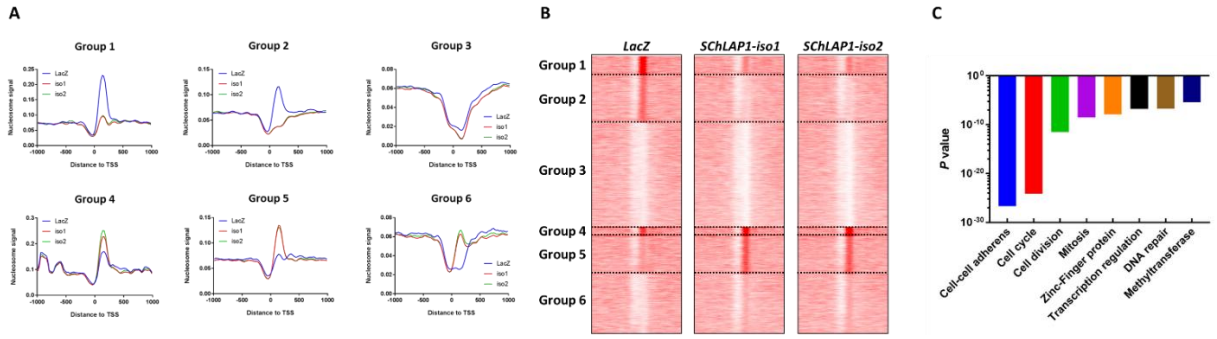


Figure 6. Nucleosome density is decreased upon SchLAP1 overexpression. A, K-means cluster of nucleosome signal (MNase-seq) in cells expressed LacZ, SchLAP isoform1 and SchLAP isoform2. Each row represents ± 1kb around the TSS of expressed genes (FPKM >1). B. The mean profiles of nucleosome signal for each subgroup. C, Gene ontology analysis of SChLAP1 regulated genes.

Task 3: To examine the effects of *SChLAP1* on H3K27Me3 levels and genome-wide binding sites of the PRC2 subunits in *SChLAP1*-expressing prostate cells.

In progress

Specific Aim-3: To determine the potential clinical utility of targeting *SChLAP1, SChLAP1*-SWI/SNF interaction and downstream histone modification as a treatment modality in metastatic prostate cancer.

Task 1: To validate the in vitro knockdown efficiency of SChLAP1 ASO in multiple prostate cancer cell lines

Based on our preliminary results, we examined the knockdown efficiency of SChLAP1 ASOs in 22RV1, LNCaP and C4-2B cells via two different ASO delivering manner (**Figure 7**).

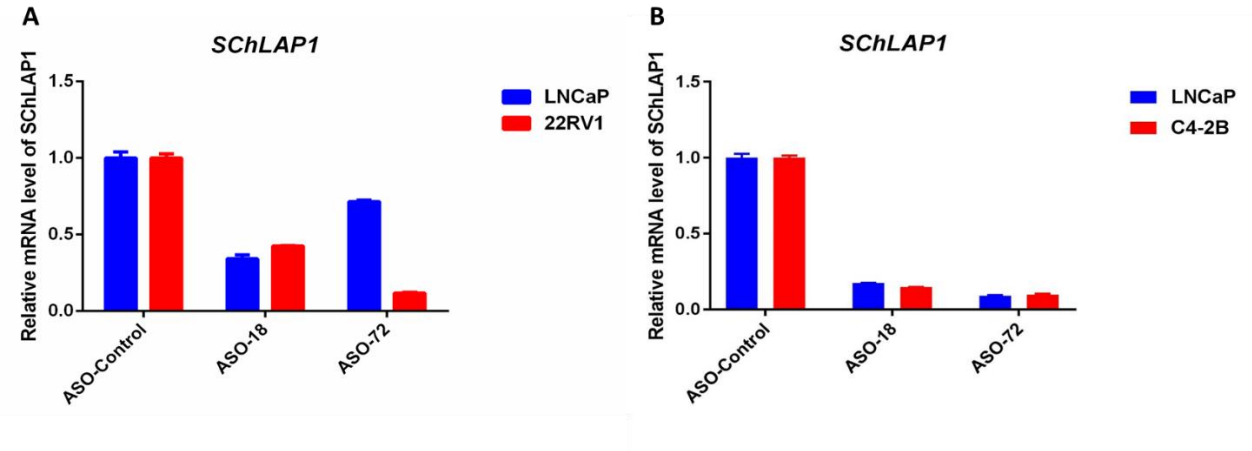


Figure 7. Validation of SChLAP1 ASO knockdown efficiency in multiple prostate cancer cell lines. A, SChLAP1 ASOs were delivered via transfection (20 nM) into LNCaP and 22RV1 cells, levels of SChLAP1 were quantified by qRT-PCR. B, SChLAP1 ASOs were delivered via free uptake (2.5 μM) into LNCaP and 22RV1 cells, levels of SChLAP1 were quantified by qRT-PCR.

Task 2: To investigate the antitumor effect of SChLAP1 ASOs in multiple prostate cancer cell lines

In LNCaP, C4-2B and 22RV1 cell, we performed *in vitro* proliferation and invasion assays on SChLAP1 knockdown cells. As shown below, knockdown of SChLAP1 didn't significantly inhibit cell proliferation (**Figure 8**), but slightly decrease invasion in LNCaP and 22RV1 cells (**Figure 9**).

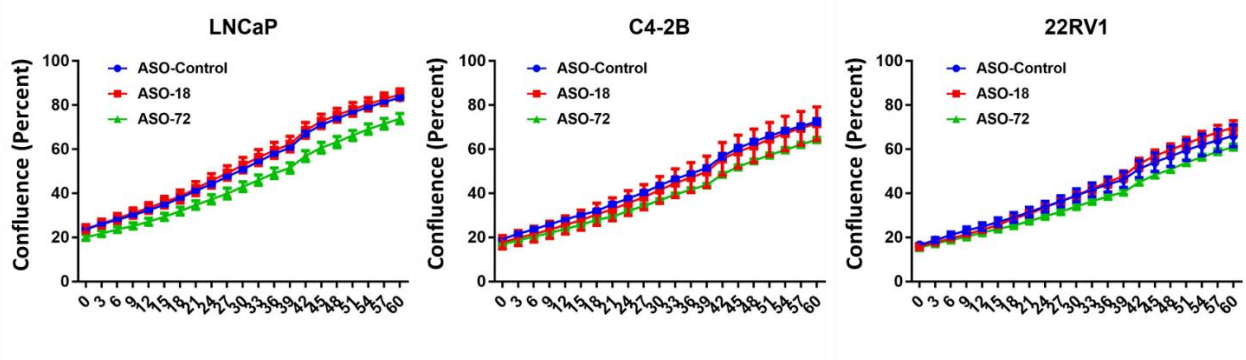


Figure 8. Proliferation of LNCaP, C4-2B and 22RV1 cells after SChLAP1 knockdown by ASOs.

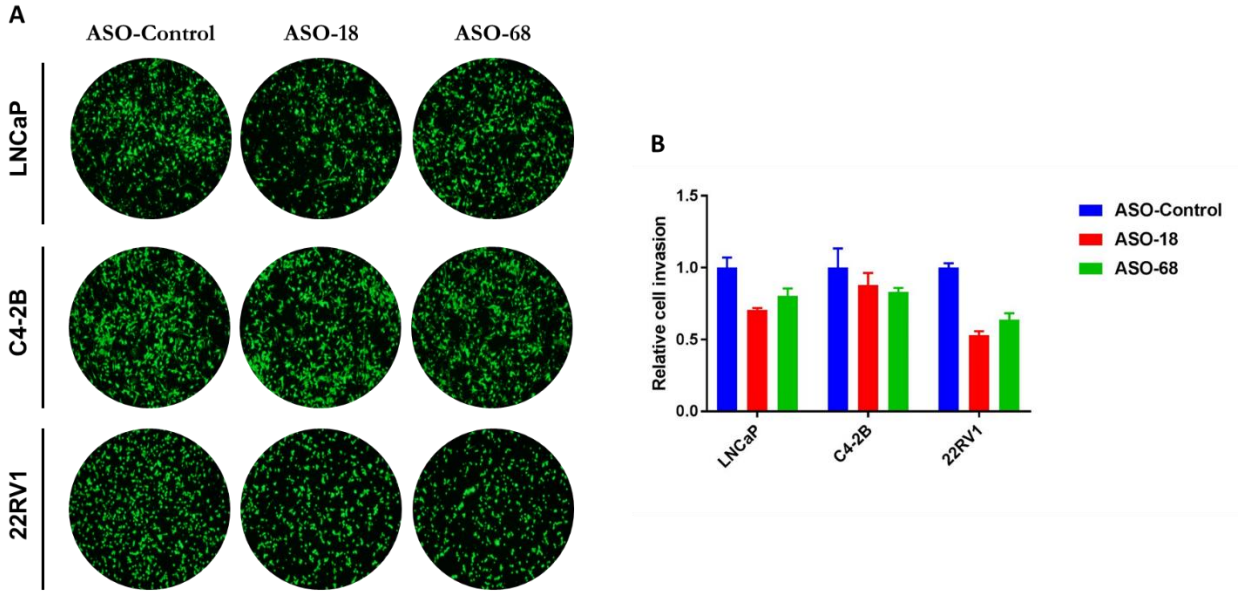


Figure 9. Invasiveness of LNCaP, C4-2B and 22RV1 cells after SChLAP1 knockdown by ASOs.

Task 3: To determine BRM as a critical synthetic lethal target in SChLAP1-epxressing prostate cancer

In progress.

Task 4: To determine the sensitivity of EZH2 inhibitor endowed by SChLAP1

We tested the growth inhibitory effect of EZH2 inhibitor (EPZ-6438) in SChLAP1 expressing prostate cancer cell lines (LNCaP and 22RV1), and SChLAP1 non-expressing prostate cancer cell line (PC3). Furthermore, IC50 of EPZ-6438 was examined in multiple SChLAP1-expressing and SChLAP1 non-expressing prostate cancer cell lines, we found EPZ-6438 is more sensitive in SChLAP1 expressing prostate cancer cells.

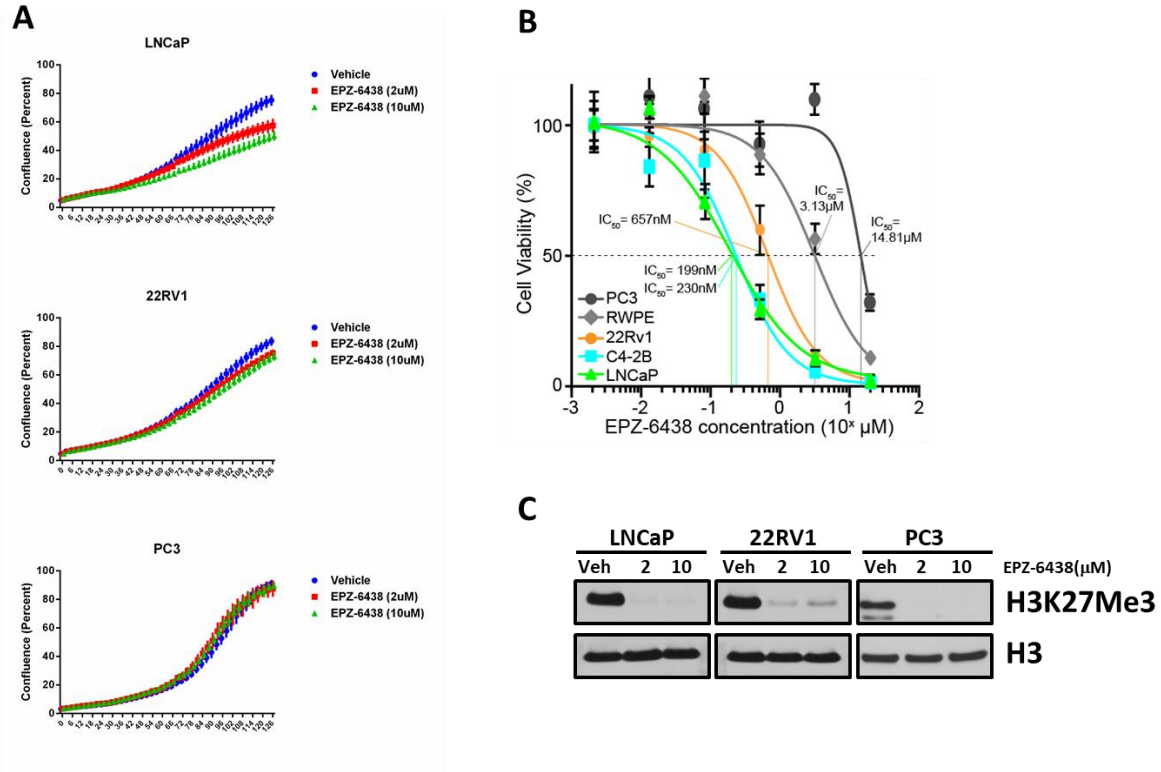


Figure 10. The anti-tumor effect of EPZ-6438 in multiple prostate cancer cell lines. A, The growth inhibitory effect of EPZ-6438 in LNCaP, 22RV1 and PC3. B, IC50 plot of EPZ-6438 in multiple prostate cancer cell lines. C, Effect of EPZ-6438 on H3K27Me3 in LNCaP, 22RV1 and PC3.

Task 5: To determine the tumor growth inhibitory effect of EPZ-6438 in C4-2B AND 22RV1 xenografts.

In C4-2B xenografts, mice dosed orally with Vehicle (0.5% CMC), Enzalutamide (30 mg/kg) or EPZ-6438 (250 mg/kg, QD or BID), respectively. Tumor volume, weight and animal body weight were monitored, and western blot was performed to check the effect of EPZ-6438 on H3K27Me3. As shown below, we found that EPZ-6438 efficiently decreased H3K27Me3 levels in C4-2B xenografts (**Figure 11D**), but the tumor volume and weight was not significantly affected by EPZ-6438 (**Figure 11A and11B**). In addition,

we tested EPZ-6438 in 22RV1 xenografts at higher dose (500 mg/kg). Like C4-2B xenografts, EPZ-6438 did not significantly inhibit 22RV1 xenografts growth *in vivo* (**Figure 11E, 11F and 11G**).

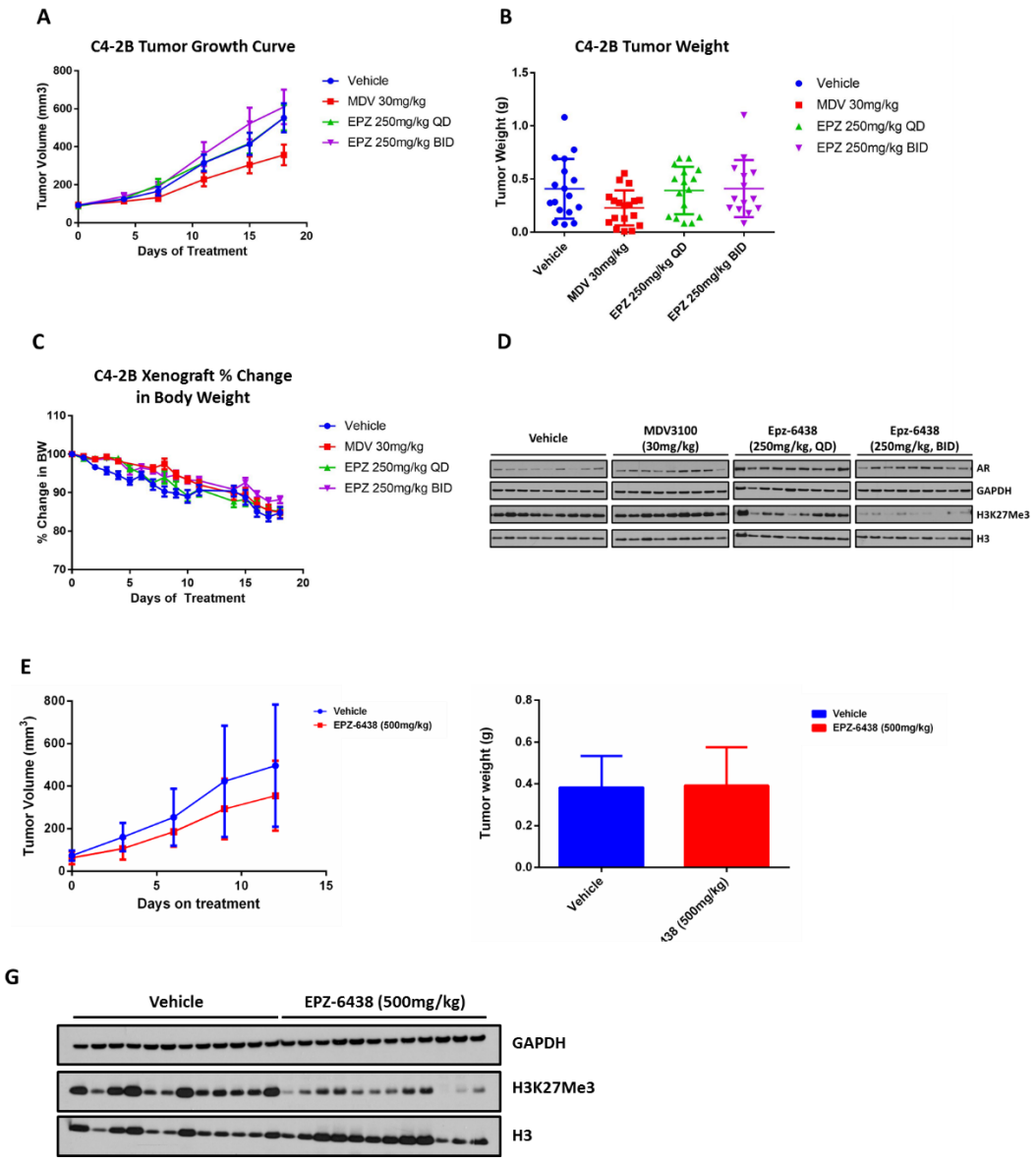


Figure 11. Tumor growth inhibitory effect of EPZ-6438 in vivo. A, B and C, Comparison of the effect of Vehicle, Enzalutamide (30 mg/kg) and EPZ-6438 (250 mg/kg, QD or BID) on tumor volume (A), tumor weight (B) and animal body weight changes (C) in C4-2B xenografts. D, The effect of EPZ-6438 on H3K27Me3 in C4-2B xenografts. E and F, The effect of Vehicle and EPZ-6438 (500 mg/kg) on tumor volume (E), tumor weight (F) in 22RV1 xenografts. G, The effect of EPZ-6438 on H3K27Me3 in 22RV1 xenografts.

What opportunities for training and professional development has the project provided?
Several opportunities for training and professional development were provided while working on this project.
One-on-one meeting with mentor: Data was discussed with mentor (Dr. Arul Chinnaiyan) in biweekly individual meetings and via monthly progress reports. Constant inputs and advice were provided.
Presentations at Scientific Conference: Data obtained was presented at scientific conferences:
1. Annual AACR conference (April, 2017)

Manuscript writing:

Weekly lab meetings: All important research findings were presented during weekly lab meetings in front of the group.

How were the results disseminated to communities of interest?
Results from the grant were published in form on a AACR poster.

What do you plan to do during the next reporting period to accomplish the goals?
By next reporting period I anticipate to accomplish all the aims proposed in the study. We also anticipate to publish manuscript describing results of co-targeting AR and EZH2 in SChLAP1-expressing prostate cancer cells.

3. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?
Once completed, the proposed project "Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA *SChLAP1*" will have substantial impact on understanding the biological function of SChLAP1 in prostate cancer progression. We have proposed to study the role of SChLAP1 in development and progression of prostate cancer as well as explore the potential of lncRNA to serve as biomarkers and therapeutic target for metastatic prostate cancer.

Till data we have discovered the potential protein binding sites of SChLAP1 and determine the specific localization pattern of SChLAP1 in prostate cancer cells and biopsies. We also showed that by interacting with SWI/SNF complex, SChLAP1 mediates significantly nucleosome occupancy changes in prostate cells. More importantly, we found tested the translational value of SChLAP1 as a biomarker and therapeutic targets for prostate cancer.

We anticipate that at the end of this proposed project, it will advance our knowledge on the critical role of lincRNA in metastatic prostate cancer, and offer potential translation opportunities for future progressive prostate cancer treatment.

What was the impact on other disciplines?

Nothing to report

What was the impact on technology transfer?

Nothing to report

What was the impact on society beyond science and technology?

Nothing to report

4. **CHANGES/PROBLEMS:**

Changes in approach and reasons for change

We proposed to systematically assess the antitumor effect of SChLAP1 ASO *in vitro* and *in vivo*, but we didn't observe significant effect of SChLAP1 on proliferation and invasion in vitro. Thus, without strong in vitro evidence, we didn't further test the anti-tumor effect of SChLAP1 ASOs in vivo.

Actual or anticipated problems or delays and actions or plans to resolve them

Since the SChLAP1 ASOs didn't exhibit strong anti-tumor effect in in vitro, we started to employ CRIPSR technique to determine the oncogenic function of SChLAP1 in LNCaP cells. As shown in **Figure 12**, we generated three monoclonal heterozygous SChLAP1 deletion lines (No.6, No.11 and No.15). Copy number of SChLAP1 in each colon is determined by smFISH (**Figure 12A**), and we found that the proliferation of all SChLAP1 loss clones are significant slower than WT clones. Next, we plan to profile the transcriptomic changes in SChLAP1 loss clones and WT clones by RNA-seq.

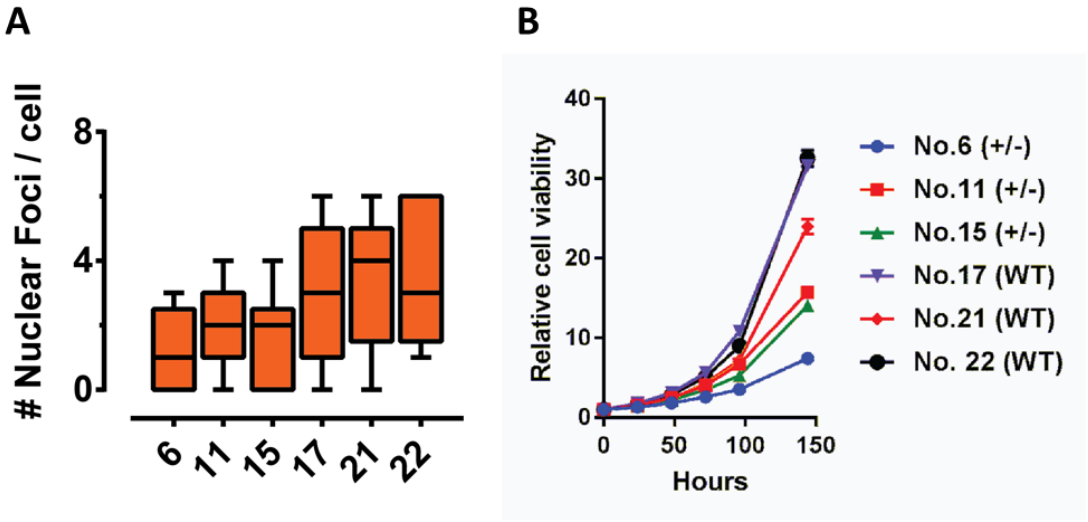


Figure 12. Loss-of-function of SChaLAP1 in LNCaP by CRIPSR technique. A, The copy number of SChLAP1 loss clones and WT clones was quantified by smFISH assay. B, The proliferation of SChLAP1 loss clones and WT clones was examined by Cell-titer Glo assay.

Changes that had a significant impact on expenditures

We performed cell proliferation and invasion assay to determine the antitumor effect of SChLAP1 ASO. We found that even through the knockdown efficiency of SChLAP1 ASOs were

similar to SChLAP1 siRNA or shRNA, but SChLAP1 ASOs didn't phenocopy the effect of SChLAP1 siRNA or shRNA on proliferation and invasion. We also proposed that SChLAP1-expressing prostate cancer cells might sensitive to EZH2 inhibitor (EPZ-6438), but our *in vitro* and *in vivo* data suggested that EPZ-6438 along didn't effectively inhibited the tumor growth. Interestingly, we found that SChLAP1-expressing prostate cancer are sensitive to the combination of EPZ-6438 and Enzalutamide, thus we will further test the co-targeting EZH2 and AR strategy in SChLAP1-expressing prostate cancer cells in vitro and in vivo.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

No Changes

Significant changes in use or care of human subjects:

None

Significant changes in use or care of vertebrate animals.

None

Significant changes in use of biohazards and/or select agents.

None

5. PRODUCTS:

Publications, conference papers, and presentations

Journal publications:

Shukla S*, Zhang X*, Niknafs YS*, **Xiao L***, Mehra R, Cieślik M, Ross A, Schaeffer E, Malik B, Guo S, Freier SM, Bui HH, Siddiqui J, Jing X, Cao X, Dhanasekaran SM, Feng FY, Chinnaiyan AM, Malik R. Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. Neoplasia. 2016 Aug;18(8):489-99.

Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y, Prensner JR, Poliakov A, Singhal U, **Xiao L**, Kregel S, Siebenaler RF, Zhao SG, Uhl M, Gawronski A, Hayes DF, Pierce LJ, Cao X, Collins C, Backofen R, Sahinalp CS, Rae JM, Chinnaiyan AM, Feng FY. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun. 2016 Sep 26;7:12791.

Other publications, conference papers, and presentations

Xiao L, Shukla S, Zhang X, Niknafs YS, Malik R, Chinnaiyan AM. Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. [abstract]. In: Proceedings of the 104th Annual Meeting of the American Association for Cancer Research; 2013 Apr 6-10; Washington, DC. Philadelphia (PA): AACR;

Technologies or techniques

Nothing to Report.

Inventions, patent applications, and/or licenses

Nothing to Report.

Other Products

Nothing to Report.

6. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Lanbo Xiao – No change. 11.4 CM

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

None

7. SPECIAL REPORTING REQUIREMENTS

None

8. **APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*

1. Shukla S, et. al., Neoplasia 2016
2. Niknafs YS, et. al., Nature Communication 2016
3. *Curriculum Vitae*_Lanbo Xiao

NEOPLASIA
www.neoplasia.com

Volume 18 Number 8 August 2016 pp. 489–499 489

Identification and Validation of *PCAT14* as Prognostic Biomarker in Prostate Cancer[1]

Sudhanshu Shukla[*,‡,2], Xiang Zhang[*,†,2], Yashar S. Niknafs[*,2], Lanbo Xiao[*,‡,2], Rohit Mehra[*,‡], Marcin Cieślik[*,‡], Ashley Ross[§], Edward Schaeffer[§], Bhavna Malik[¶], Shuling Guo[#], Susan M. Freier[#], Huynh-Hoa Bui[#], Javed Siddiqui[*,‡], Xiaojun Jing[*,‡], Xuhong Cao[*,‡], Saravana M. Dhanasekaran[*,‡], Felix Y. Feng[*,¶,3], Arul M. Chinnaiyan[*,‡,,††,‡‡,3] and Rohit Malik[*,‡,3]**

*Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, USA; [†]Department of Urology, Qilu Hospital of Shandong University, Jinan, China; [‡]Department of Pathology, University of Michigan, Ann Arbor, USA; [§]James Buchanan Brady Urological Institute, Johns Hopkins University, USA; [¶]Department of Radiation Oncology, University of Michigan, Ann Arbor, USA; [#]Ionis Pharmaceutical, CA; [**]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, USA; [††]Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA; [‡‡]Comprehensive Cancer Center, University of Michigan, Ann Arbor, USA

Abstract

Rapid advances in the discovery of long noncoding RNAs (lncRNAs) have identified lineage- and cancer-specific biomarkers that may be relevant in the clinical management of prostate cancer (PCa). Here we assembled and analyzed a large RNA-seq dataset, from 585 patient samples, including benign prostate tissue and both localized and metastatic PCa to discover and validate differentially expressed genes associated with disease aggressiveness. We performed Sample Set Enrichment Analysis (SSEA) and identified genes associated with low versus high Gleason score in the RNA-seq database. Comparing Gleason 6 versus 9+ PCa samples, we identified 99 differentially expressed genes with variable association to Gleason grade as well as robust expression in prostate cancer. The top-ranked novel lncRNA *PCAT14,* exhibits both cancer and lineage specificity. On multivariate analysis, low *PCAT14* expression independently predicts for BPFS ($P = .00126$), PSS ($P = .0385$), and MFS ($P = .000609$), with trends for OS as well ($P = .056$). An RNA in-situ hybridization (ISH) assay for *PCAT14* distinguished benign vs malignant cases, as well as high vs low Gleason disease. *PCAT14* is transcriptionally regulated by AR, and endogenous *PCAT14* overexpression suppresses cell invasion. Thus, Using RNA-sequencing data we identify *PCAT14,* a novel prostate cancer and lineage-specific lncRNA. *PCAT14* is highly expressed in low grade disease and loss of *PCAT14* predicts for disease aggressiveness and recurrence.

Neoplasia (2016) 18, 489–499

Address all correspondence to: Arul M. Chinnaiyan, M.D., Ph.D., Investigator, Howard Hughes Medical Institute, American Cancer Society Professor, S. P. Hicks Endowed Professor of Pathology, Professor of Pathology and Urology, Comprehensive Cancer Center, University of Michigan Medical School, 1400 E. Medical Center Dr. 5316 CCGC, Ann Arbor, MI 48109-0602.
E-mail: arul@med.umich.edu

[1]Funding/Support: This work was supported in part by the Prostate Cancer Foundation (F.Y.F, A.M.C), the National Institutes of Health Prostate SPORE (P50CA186786 to A.M.C) and the Early Detection Research Network (U01CA111275 and U01CA113913 to A.M.C). A.M.C. is supported by the Alfred A. Taubman Institute, Howard Hughes Medical Institute and the American Cancer Society. R.M is supported by a Department of Defense postdoctoral award (W81XWH-13-1-0284). R.M. and M.C. are supported by a Prostate Cancer Foundation Young Investigator award. The sponsors played no role in the design and conduct of the study.
[2]Equal Contributions.
[3]Co-senior authors.
Received 5 July 2016; Accepted 6 July 2016

1476-5586
http://dx.doi.org/10.1016/j.neo.2016.07.001

Introduction

Early detection of prostate cancer, largely facilitated by the advent of PSA screening, has also been attributed to over-diagnosis and overtreatment of this disease [1–3]. While coupling PSA screening with other biomarkers such as the long non-coding RNA (lncRNA) transcript *PCA3* or gene fusions events (such as TMPRSS2-ERG) have increased specificity of cancer diagnosis, these biomarkers have limited utility in stratifying patients in terms of prognosis [4,5]. While stratifying patients into risk groups based on clinicopathologic features is currently used to guide treatment decisions [6], it is clear that current stratification approaches need to be further refined to allow better personalization of therapy. Thus, identifying molecular biomarkers to distinguish indolent versus aggressive disease would address an unmet need in the clinical management of prostate cancer.

Advances in next-generation sequencing technologies have enabled thorough characterization of cancer transcriptomes, especially in unraveling the realm of non-coding RNAs (ncRNAs) [7,8]. In particular, lncRNAs, a class of ncRNAs, have gained increasing attention as biomarkers due to their tissue- and cancer-specific expression profile [9]. In this study, we assembled and analyzed a large RNA-seq compendium compiled from recent publications from consortiums such as The Cancer Genome Atlas (TCGA), the Prostate Cancer Foundation/Stand Up to Cancer international team, and others to identify differentially expressed genes (both protein coding and non-coding genes), that are associated with indolent versus aggressive disease [10,11]. Our results identify *PCAT14*, a prostate cancer- and lineage-specific lncRNA, as a top differentially expressed gene in this context. We characterize *PCAT14* preclinically and demonstrate that it correlates inversely in expression with disease aggressiveness and adds to conventional clinicopathologic risk factors in predicting prognosis in prostate cancer patients. Finally, we develop a novel in-situ hybiridation (ISH)-based approach for detecting *PCAT14* in clinical samples.

Material and Methods

RNA-Seq Data Set

Prostate RNA-seq cohort (n = 585) containing 52 benign prostate tissues, 501 primary prostate cancers, and 132 metastatic prostate cancers was used in this study. For nomination of Gleason associated genes, we compared low Gleason tumors (Gleason 6, n = 45) to high Gleason tumors (Gleason 9+, n = 140).

RNA-seq Data Processing

TCGA prostate Fastq files were obtained from the CGhub. Reads were aligned using STAR version 2.4.2 [12] and read abundance was calculated using FeatureCounts version 1.4.6 [13].

RNA-Seq Differential Expression Testing

Differential expression testing was performed using the Sample Set Enrichment Analysis (SSEA) tool described previously [7]. Briefly, following count data normalization, SSEA performs the weighted KS-test procedure described in GSEA [14]. The resulting enrichment score (ES) statistic describes the enrichment of the sample set among all samples being tested. To test for significance, SSEA enrichment tests are performed following random shuffling of the sample labels. These shuffled enrichment tests are used to derive a set of null enrichment scores (1000 null enrichment scores computed). The nominal p value reported is the relative rank of the observed enrichment score within the null enrichment scores. Multiple hypothesis testing is performed by comparing the enrichment score of the test to the null normalized enrichment score (NES) distributions for all transcripts in a sample set. This null NES distribution is used to compute FDR q values in the same manner used by GSEA [14]. SSEA percentile score determined by ranking the genes in each analysis by their NES score.

Tissue Expression Heatmap Generation

The "gplots" R-package was used to generate heatmaps using the *heatmap.2* function. Expression was normalized as log2 of the fold-change over the median of the normal samples for each transcript. Unsupervised hierarchical clustering was performed with the *hclust* function, using Pearson correlation as the clustering distance, using the "ward" agglomeration method.

Identification of Genes Differentially Expressed in Prostate Cancer of Varying Gleason Score

Differentially expressed Gleason associated genes were identified as any gene with an SSEA FDR< 0.01 when comparing Gleason 6 primary tumors to Gleason 9+ primary tumors. Filtering for expression levels in tissues was done by enforcing that each gene had >5FPKM expression in the top 5% of prostate tumor samples. Filtering for overexpression in cancers versus normal was done by enforcing an SSEA FDR of <0.0001 in an analysis comparing the TCGA prostate cancer vs normal tissues. Tissue specificity percentile was determined as the SSEA percentile for each gene in an SSEA analysis comparing the TCGA prostate samples to all other TCGA tumors in our multi-tissue compendium [7].

Clinical Analysis

To assess the prognostic value of *PCAT14*, microarray data was obtained from the Johns Hopkins University (JHU) (N = 355). Patients were treated with prostatectomy and subsequently received no adjuvant or salvage treatment until metastasis. Microarray processing and normalization was performed as described previously [15]. *PCAT14* expression was calculated by taking the mean expression of probe sets mapping to exons. High/low *PCAT14* was determined by splitting on the median expression level. Kaplan–Meier curves are shown and statistical inference was performed using the Log-rank test. Multivariate analysis was performed using Cox regression. Age was treated as a continuous variable. PSA was grouped into low (<10 ng/ml), intermediate (10–20 ng/ml), and high (>20 ng/ml). Surgical margin status (SMS), seminal vesicle invasion (SVI), extracapsular extension (ECE), and lymph node invasion (LNI) were treated as binary variables. Gleason score was grouped into low (≤7) or high (8–10). Association of *PCAT14* and clinicopathologic variables was evaluated using a *t*-test for continuous variables, and a chi-squared test for categorical variables. Statistical significance was set as a two-sided *p*-value <0.05. All analyses were performed in R 3.1.2.

ISH Analysis

PCAT14 ISH was performed on thin (approximately 4 μm thick) TMA sections (Advanced Cell Diagnostics, Inc., Hayward, CA), as described previously [16,17]; in parallel, *PCAT14* ISH was performed on previously identified positive and negative control index formalin-fixed paraffin embedded (FFPE) tissue sections. All slides were examined for *PCAT14* ISH signals in morphologically intact cells and scored manually by a study pathologist (Rohit Mehra). Specific *PCAT14* ISH signal was identified as brown, punctate dots, and expression level was scored as follows: 0 = no staining or less than 1 dot per 10 cells, 1 = 1 to 3 dots per cell, 2 = 4 to 9 dots per cell (few

or no dot clusters), 3 = 10 to 14 dots per cell (less than 10% in dot clusters), and 4 = greater than 15 dots per cell (more than 10% in dot clusters). For each evaluable tissue core, a cumulative ISH product score was calculated as the sum of the individual products of the expression level (0 to 4) and percentage of cells (0 to 100) (i.e., [A% × 0] + [B% × 1] + [C% × 2] + [D% × 3] + [E% × 4]; total range = 0 to 400). For each tissue sample, the ISH product score was averaged across evaluable TMA tissue cores. All quantitative data were shown as mean ± S.D. To obtain significance in the difference between two groups was performed by two-sided t test using Graph Pad Prism 6.02 software.

Cell Lines, Tissues and Reagents

All prostate cell lines used in this study were purchased from the American Type Culture Collection (ATCC), cultured according to their recommendations and were periodically checked for mycoplasma contamination and genotyped to confirm identity. For androgen treatment experiments, VCaP cells were pre-cultured in androgen-free charcoal-stripped medium for 48 hours and treated with 10 nM dihydrotestosterone (DHT) or 10 μM MDV3100 or vehicle (ethanol) for indicated time points before cells were harvested for RNA isolation. For drug treatment experiments, LNCaP cells were treated with the 5–20 μM DNA methylation inhibitor 5-aza-2′-deoxycytidine (5-aza) (catalog: A3656-5MG, Sigma), or DMSO for 5 days. RNA was isolated 24 h after drug treatment and expression was analyzed by qRT-PCR.

Prostate specimens were acquired from the patients who underwent radical prostatectomy and from the Rapid Autopsy Program at the tissue core of University of Michigan as part of the University of Michigan Prostate Cancer Specialized Program Of Research Excellence (S.P.O.R.E.). Informed consents were obtained from each patient.

RNA Isolation and qPCR Analysis

Total RNA was extracted using Trizol reagent and an RNeasy Micro Kit (Qiagen) with DNase I digestion according to the manufacturer's protocols. RT-PCR was performed from total RNA using Superscript III (Invitrogen) with random primers (Invitrogen). Quantitative PCR (qPCR) was performed using Fast SYBR Green Master Mix (Applied Biosystems) on a 7900HT Fast Real-Time PCR system (Applied Biosystems). All oligonucleotide primers were purchased from Integrated DNA Technologies (Coralville, IA) are sequence of each primer is listed in Supplementary Table 4. Primer specificity was determined by sequence verifying the PCR products using the University of Michigan Sequencing Core facility.

Rapid Amplification of cDNA Ends (RACE)

5′ and 3′ RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer's protocols. RACE PCR products obtained using Platinum Taq high-fidelity polymerase (Invitrogen), were resolved on a 1.5% agarose gel. Individual bands were gel purified using a Gel Extraction kit (Qiagen), and cloned into PCR4 TOPO vector, and sequenced using M13 primers.

Knock Down Studies

MDA-PCa-2b and VCaP cells were seeded in biocoated 6-well plates at 60% confluency, incubated overnight, and transfected with 50 nM siRNAs targeting different exons of *PCAT14* or non-targeting siRNAs, using RNAi MAX reagent (Invitrogen) per manufacturer's instructions. RNA was harvested 48 h after transfection. Functional experiments were performed at indicated time points. Sequence of all the siRNA used in shown in Supplementary Table 4.

Nuclear-Cytoplasmic Subcellular Fractionation

Nuclear-cytoplasmic fraction of MDA-PCa-2b and VCaP cells was performed using an NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Scientific) following manufacturer's instructions, followed by RNA isolation and qPCR analysis.

CRISPR Based Overexpression of PCAT14

Stable cell lines overexpressing *PCAT14* endogenously were made using previously published protocol [18]. Briefly, guide RNAs targeting promoter region of *PCAT14* (Supplementary Table 4) were designed using online tool at http://crispr.mit.edu/ and cloned into sgRNA-MS2 vector using lenti sgRNA(MS2) zeo backbone. Lentiviral particles expressing *PCAT14* sgRNA-MS2 were generated by the University of Michigan vector core. To generate LNCaP or PC3 cell over expressing *PCAT14*, first cells were seeded into 100 mm dish and transduced with Lenti dCAS-VP64 (blasticidin) and Lenti-MS2-p65-HSF1 (hygromycin) vectors. After 2 days, cells were selected with 4 μg/ml Blasticidin and 200 μg/ml Hygromycin. Cells stably expressing dCAS-VP64 and MS2-p65-HSF1 cells were then seeded in 6-well plates and infected with *PCAT14* sgRNA-MS2 lentivirus. 24 hours later, cells were selected with triple antibiotics: 4 μg/ml Blasticidin, 200 μg/ml Hygromycin and 800 μg/ml Zeomycin for 1 week. Expression of *PCAT14* in these cells was verified by qPCR.

In Vitro FluoroBlok Tumor Invasion Assay

The In vitro FluoroBlok Tumor Invasion Assay (BD) was performed as previously described [19]. Briefly, after rehydration of the BD FluoroBlok membrane, 500 ul of serum-free RPMI medium resuspended prostate cancer cells (PC3, 50,000 cells per well, or LNCaP, 100,000 cells per well) were seeded into the apical chambers. 750 ul RPMI medium containing 10% FBS were added to the basal chamber as chemoattractant. Then plates were incubated at 37 °C, 5% CO_2 for 24 hours. Following incubation, medium from the apical chambers were removed, and the inserts were transferred to a 24-well plate containing 500ul/well of 4ug/mL Calcein AM (Invitrogen) in Hanks buffered saline. Plates were incubated for 1 hour at 37 °C, 5% CO_2, then pictures of invaded cells were taken by using inverted fluorescence microscope (Olympus), and quantified by ImageJ software [20].

Oncomine Concepts Analysis of the PCAT14 Signature

Gene that positively correlated ($R^2 > 0.35$, n = 591) with *PCAT14* in TCGA RNA-seq data were selected and uploaded into Oncomine database [21] as custom concepts (Supplementary Table 2). All the prostate cancer concepts with odds ratio> 2.0 and *p*-value $<1 \times 10^{-4}$ were selected. For simplicity, top 4 concepts (based on odds ratios) were selected for representation. We exported these results as the nodes and edges of a concept association network and visualized the network using Cytoscape version 3.3.0. Node positions were computed using the Edge-weighted force directed layout in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.

Statistics

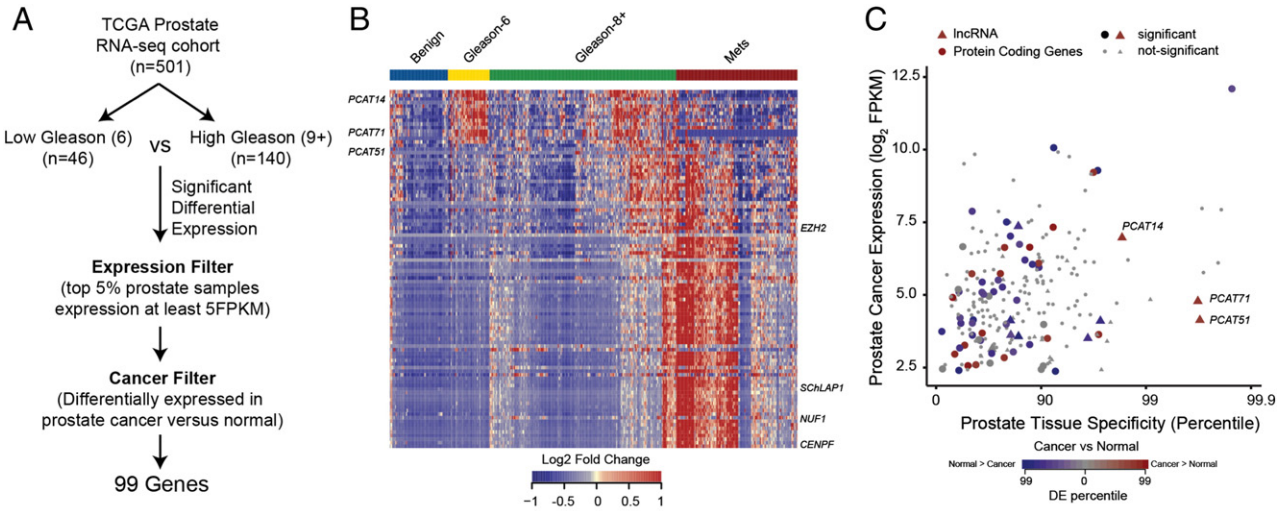
All quantitative data were shown as mean ± S.D. To obtain significance in the difference between two groups was performed by two-sided t test or ANOVA using Graph Pad Prism 6.02 software.

Results

Identification of Genes Associated With Gleason Grade in Prostate Cancer

Comprehensive molecular characterization of common cancer types has become feasible with the recent availability of large next generation sequencing datasets on tumor tissues. To identify genes (both coding and non-coding) associated with aggressive prostate

cancer, we assembled a large prostate RNA-seq cohort (n = 585) containing 52 benign prostate tissues, 501 primary prostate cancers, and 132 metastatic prostate cancers. We performed differential expression testing utilizing a non-parametric tool we developed for RNA-seq data called Sample Set Enrichment Analysis [7]. In order to nominate the most intriguing biomarkers associated with aggressive disease, we compared low Gleason tumors (Gleason 6, n = 45) to high Gleason tumors (Gleason 9+, n = 140) and applied filters for substantial expression in prostate tumor tissue (>5PKM in the top 5% of prostate samples), and significant differential expression in



D

Gene Name	Category	Gleason 9+ Percentile	Prostate Expression (95%ile FPKM)	Cancer vs Normal Percentile	Tissue Sepcificity Percentile
PCAT14	lncRNA	99.57	125.589	97.93	98.40
TSPAN1	Protein coding	99.63	594.854	98.05	96.92
PTRHD1	Protein coding	99.38	160.925	99.03	92.41
DUS1L	Protein coding	99.57	99.683	99.02	87.25
GMDS	Protein coding	99.91	99.117	98.31	77.82

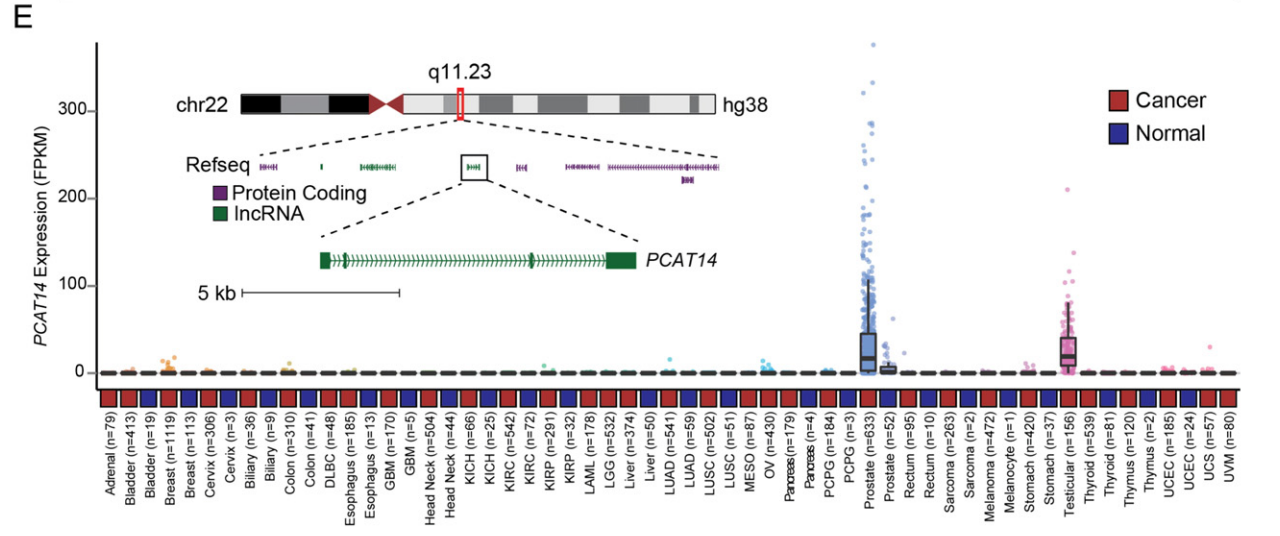


Figure 1. Identification of lncRNA *PCAT14* as a prostate cancer biomarker. A. Schematic representation of the workflow utilized to identify highly-expressed, prostate cancer specific genes associated with low-Gleason disease. B. Heatmap depiction of the lncRNA and protein coding genes differentially expressed (n = 99) between Gleason 6 versus 9+ analysis in TCGA prostate RNA-seq data. Relative expression of these genes in benign and metastatic prostate cancer tissues [11] are also displayed alongside for comparison. Expression is depicted as log2 of the fold-change over the median of the Gleason 6 samples for each gene. Patients grouped by cancer progression/Gleason score. Rows represent genes and columns represent samples. C. Scatterplot showing the expression level, prostate tissue specificity, and prostate cancer association of protein coding (solid circle) and lncRNA (solid triangles) genes identified in 1 A. Expression is represented by the FPKM value for the 95th percentile prostate cancer sample. Cancer versus normal and prostate tissue specificity are represented by the percentile score for each gene in an SSEA analysis. D. The top five, Gleason 6 associated genes listed in the order of prostate tissue specificity. E. Expression of *PCAT14* across all cancer and normal tissue type represented in the TCGA. Inset shows genome browser view of *PCAT14* genomic location.

prostate cancers versus normal (SSEA, FDR <0.0001) leaving a total of 99 candidates genes (Figure 1*A*, Supplementary Table 1). Interestingly, clustering analysis revealed signature expression patterns, specifically associated with low, high Gleason and metastatic status and included both novel and previously characterized genes (Figure 1*B*). *CENPF* and EZH2, protein coding genes with a known association with high grade prostate cancer were rediscovered through this analysis [22,23]. Similarly, we rediscovered *SChLAP1* a long non-coding RNA (lncRNA) associated with aggressive prostate cancers [15,17] in our analysis (Figure 1*B*). With the goal of identifying potential biomarkers that distinguish indolent prostate cancers, we focused on genes enriched in low grade disease that are expressed highly in prostate tissue and that also show prostate cancer and tissue specificity (Figure 1*C*). Interestingly, a lncRNA, *PCAT14* appeared to be one of the top low-Gleason-associated genes with robust prostate tissue expression, substantial prostate tissue specificity, and significant overexpression in prostate cancers versus normal (Figure 1*D*). In fact, among all genes (coding and non-coding), *PCAT14* ranked among the top 5 in terms of expression level, Gleason 6 versus 9+ association, and cancer versus normal association (Figure 1*D*). Additionally, among the top 5 candidate genes, *PCAT14* was the only gene to exhibit striking prostate tissue specificity, a particularly relevant metric for a potential biomarker (Figure 1*E*). The remaining 4 genes exhibited variable prostate tissue specificity (Supplementary Figure 1). *PCAT14* is a poly-exonic gene found within a gene desert on chromosome 22, with a striking prostate cancer and lineage specific expression pattern across the >10,000 TCGA cancer and normal tissue samples (Figure 1*E*). For these reasons, we elected to pursue *PCAT14* as a promising biomarker that can identify low grade prostate cancer.

Genomic Organization and Regulation of PCAT14

We collected multiple lines of evidence from both experimental data and available annotations to consolidate the genomic organization of *PCAT14*. Based on assembled reads from RNA-seq data assembled in the MiTranscriptome [7], we predicted the structure of the *PCAT14* transcript variants (Supplementary Figure 1A). Additionally, as an independent approach to define the exon structure of *PCAT14,* we performed rapid amplification of cDNA ends (RACE) in two prostate cancer cell lines VCaP and MDA-PCa-2b that express *PCAT14* at high levels (Supplementary Figure 1B and C). Our analyses show that the *PCAT14* gene is located on chr22-q11.2 and contains 4 exons. Among the four transcript isoforms, the 2.3 kb variant-1 demonstrates the highest expression (Supplementary Figure 1D). Next, using published ChIP-seq data in VCaP cells [24], we show that *PCAT14* has all the histone marks (H3K4me3, H3K36me3, H3K27ac) associated with actively transcribed genes (Figure 2A). We further performed subcellular fractionation followed by qPCR to show that *PCAT14* is distributed equally between nuclear and cytoplasmic compartments (Figure 2*B*).

Androgen receptor plays a major role To identify any potential regulation of *PCAT14* gene by androgen, we assessed the presence of AR peaks in *PCAT14* genomic region using AR-ChIP-seq data generated in VCaP cells [24] and saw significant AR peaks in *PCAT14* loci. Some of these peaks were also enhanced upon treatment with DHT and were suppressed upon treatment with AR antagonist MDV3100 or bicalutamide (Figure 2*C*). To corroborate this finding, we assessed the expression of *PCAT14* mRNA in VCaP cells upon AR stimulation. Similar to the canonical AR targets such as *KLK3* and *FKBP5*, *PCAT14* expression was also significantly elevated (four fold in 24 hours) upon DHT stimulation (Figure 2*D*) and suppressed by MDV3100 treatment (Figure 2*E*). In another line of investigation, we queried if epigenetic regulation might play a role in the prostate cancer and lineage specific expression of *PCAT14* observed in tissue samples (Figure 1*E*). Using a prostate cancer cell line (LNCaP) model we show significant elevation of *PCAT14* expression when treated with 5-azacytidine (5-Aza), a DNA demethylation agent, suggesting a potential role for promoter methylation in regulation of *PCAT14* (Figure 2*F*). However, our attempt to capture this event in TCGA tissue samples where Infinium 450 K DNA methylation array data is available was inconclusive, due to the lack of probes in *PCAT14* promoter region. Taken together we show *PCAT14* is an AR target gene that may also be subjected to epigenetic regulation in prostate cancer.

Clinical Association of PCAT14

Having observed an inverse correlation of *PCAT14* with Gleason Score (GS) in our RNA-seq cohort, we next assessed the association of *PCAT14* expression with clinical outcomes in prostate cancer. For this analysis we first divided samples into 7 groups (benign, GS-6, GS-7 (3 + 4), GS7 (4 + 3), GS-8, GS-9 and Mets) and examined the expression of *PCAT14* using two different datasets (TCGA and Taylor et al.). We identified a significant decrease in *PCAT14* expression as Gleason grade increased in both cohorts (Figure 3*A* and *B*). Importantly, in the large TCGA dataset, expression was significantly different between GS6 and all other groups except GS7 (3 + 4). We next assessed the diagnostic ability of *PCAT14* to identify prostate cancers versus normal. In both the TCGA and Taylor prostate cancer cohorts, *PCAT14* expression was able to significantly distinguish cancer from normal with an AUC of 0.837 and 0.823 respectively (Figure 3*C*) supporting its utility as a diagnostic biomarker.

Using an alternate approach to further characterize the clinical associations of *PCAT14*, we performed a "guilt-by-association" analysis, assessing the clinical significance of the protein-coding genes most correlated with *PCAT14* (Supplementary Table 2) in the TCGA prostate cancer cohort, leveraging cancer microarray data from the Oncomine resource [21]. As expected, genes positively correlated with *PCAT14* were upregulated in cancer vs normal analysis and were downregulated in clinically advanced prostate cancer (Figure 3*D*). Interestingly, we found a striking association of *PCAT14* correlated genes with concepts related to better prognosis (Figure 3*D*), and these genes were under-expressed in recurrent and hormone refractory prostate cancer suggesting that *PCAT14* may be a marker of better clinical outcomes in prostate cancer. In contrast, genes that positively correlated with *SChLAP1*, a lncRNA known to be associated with clinically aggressive prostate cancer, were found to be overexpressed in advanced prostate cancer as well as in cancer with poor outcomes [15,17].

To further investigate the association of *PCAT14* with favorable clinical outcomes in prostate cancer, we performed Cox regression analysis on a cohort of 355 patients (John Hopkins University (JHU) cohort) who did not receive treatment prior to metastasis (median follow-up 9 years). Univariate analysis showed that, patients with high *PCAT14* expression were significantly associated with better BPFS (P = .000062; HR = 0.59 [0.45–0.76]), MFS (P = .00016; HR = 0.46 [0.32–0.66]), PSS (P = .0067; HR = 0.47[0.27–0.82]) and OS (P = .022; HR = 0.57 [0.35–0.93]) (Figure 4*A-D*). In a Cox multivariate analysis including clinicopathologic variables, *PCAT14* stands out as a significant independent predictor of PSS (P = .0385; HR = 0.55 [0.31–0.97]), MFS (P = .000609; HR = 0.52[0.36–0.76])

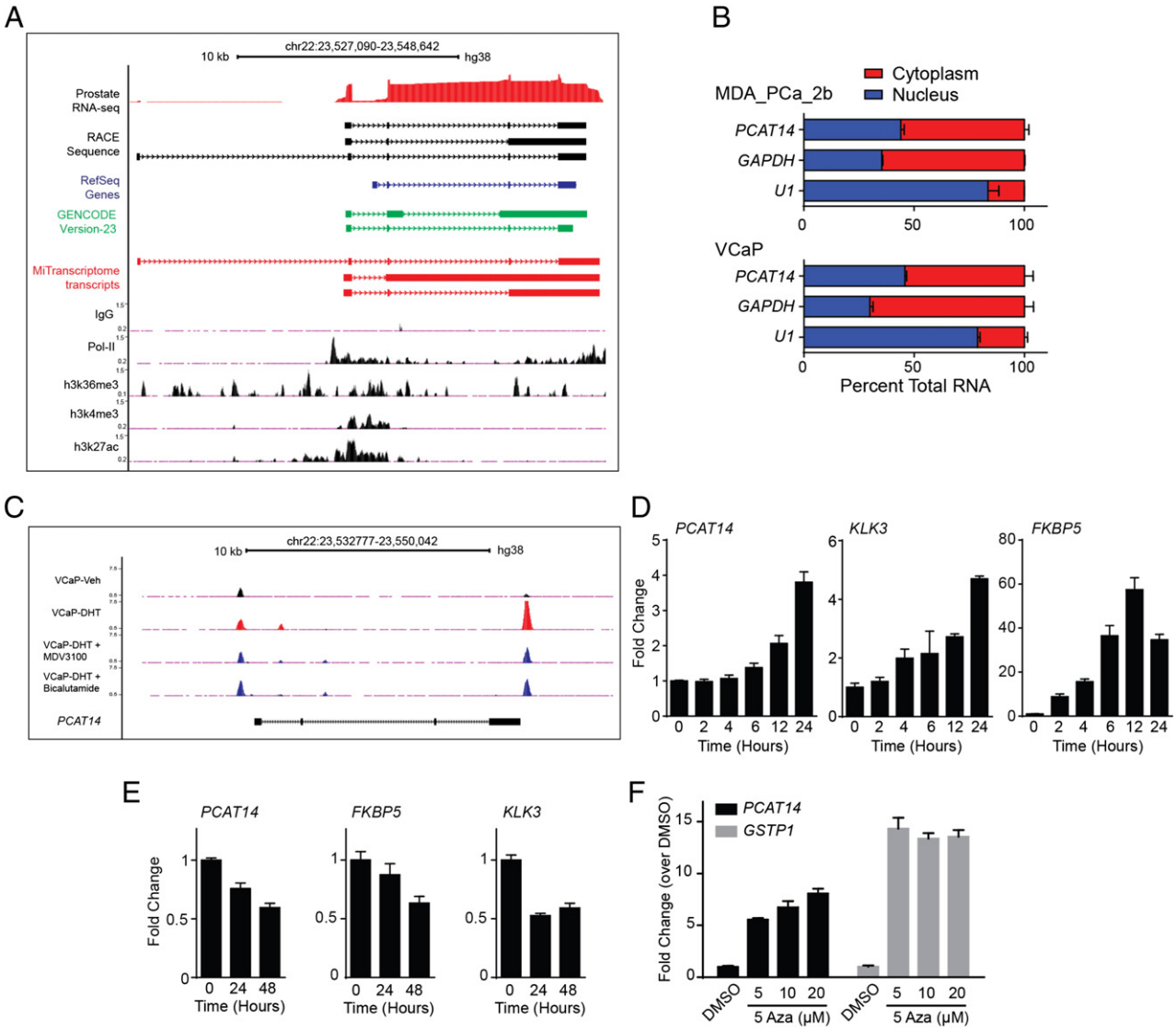


Figure 2. Subcellular localization and regulation of *PCAT14*. A. Genome browser view of *PCAT14* locus. ChIP-seq tracks for H3K4me3, H3K27ac, H3K36me3 and Pol-II generated in prostate cancer VCaP cells are shown. Prostate RNA-seq reads, transcript schematic based on RACE results and Refseq, GENCODE, MiTranscriptome assembly annotations are also provided. Solid blocks indicate exons while thin lines intron and arrows indicate the genomic orientation. B. Bar plots represent the subcellular localization of *PCAT14* in prostate cell lines. *PCAT14* transcript was equally found in both cytoplasmic (red) and nuclear (blue) compartments in both MDA-PCa-2b and VCaP cell lines. GAPDH and U1 RNA were used as controls. C. Genome browser view of the *PCAT14* genomic locus for AR ChIP-seq data tracks obtained from VCaP cells treated with either vehicle (black) or dihydrotestosterone (DHT) alone (Red) or combinations (dark blue) including DHT + MDV3100 and DHT + Bicalutamide. Significant AR binding observed in each data track are represented as peaks. D-E. Histograms represent the expression of *PCAT14*, *TMPRSS2* and *KLK3* in VCaP cells after treatment with 10 nM DHT or with MDV3100 for indicated time points. F. Bar plots represent re-expression of *PCAT14* and *GSTP1* in LNCaP cells after treatment with 5-Aza deoxycytidine (5-Aza) for 5 days at indicated concentrations.

and BRFS (P = .00126, HR = 0.64 [0.49–0.84]), with borderline significance for OS (Table 1, Supplementary Table 3). In addition, we also analyzed the association of *PCAT14* expression with clinical outcome in two independent data sets of 140 (Taylor et al) and 377 (TCGA) patients using the statistical approaches mentioned above [25]. Similar to JHU cohort, high *PCAT14* expression predicted for better BRFS (Figure 4*E*) and MFS (Figure 4*F*). We also show that high PCAT14 expression was predictor of better prognosis in lower Gleason grade samples (Supplementary Figure 3B).

PCAT14 Expression In-Situ

LncRNA detection in cancer tissue sections by RNA in-situ hybridization (RNA-ISH) technology has similar clinical utility as immunohistochemical evaluation of protein biomarkers [16,26]. Hence we evaluated *PCAT14* transcript levels in PCa FFPE tissues using specific probes to perform a RNA-ISH. We first probed a panel of FFPE sections derived from either murine prostate, kidney, lung (negative controls) or xenografts from MDA-PCa-2b cells, a cell line that expresses *PCAT14* at high levels (positive control). As expected,

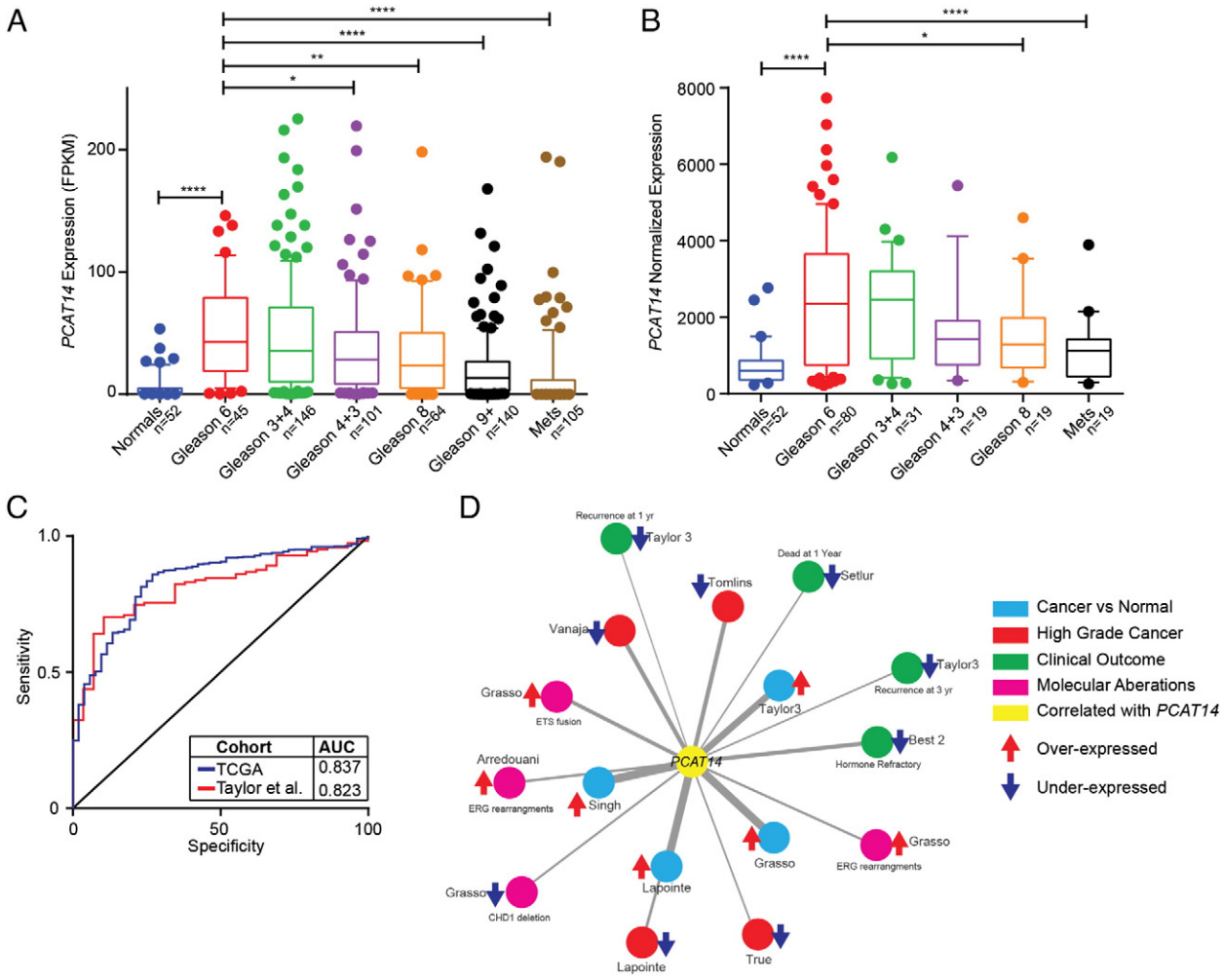


Figure 3. *PCAT14* is marker of low grade tumros. **A-B**. Expression of *PCAT14* in samples distinguished by Gleason grade in TCGA (A), Taylor (B) cohorts. (* = $P < .05$, ** = $P < .01$, **** = $P < .0001$; compared to Gleason 6). C. ROC analysis of *PCAT14* expression in the TCGA and Taylor cohorts. D. Network representation of genes positively correlated with *PCAT14* in localized prostate cancers using Oncomine concepts analysis and visualized with the Force-Directed Layout algorithm in the Cytoscape tool [29]. Node names are assigned according to the author of the primary study [25,30–38]. Nodes are colored according to the concept categories indicated in the figure legend. Thickness of the edges implies higher odds ratio.

high levels of specific signal was present in MDA-PCa-2b xenografts while no expression/staining was seen in the negative control murine tissues (Supplementary Figure 5A and B). Consistent with the cell fractionation data, expression of *PCAT14* was seen in both nuclear and cytoplasmic compartments. Next we obtained frozen and matched formalin fixed paraffin embedded (FFPE) tissues sections derived from a patient radical prostatectomy specimen with Gleason score 3 + 3 = 6 disease. q-PCR analysis on cDNA from frozen tissues derived from this specimen shows a 7–8 fold increase in *PCAT14* expression in cancer compared to the adjacent benign tissue (Figure 5*A*). RNA-ISH also demonstrated that *PCAT14* is differentially expressed in PCa as we saw striking difference of transcript expression with high signals located in the prostatic adenocarcinoma glands and with no/minimum staining in the benign section (Figure 5*B*). To further expand these results, we performed RNA-ISH on a PCa tissue microarray (TMA, n = 129) (Figure 5*C*) and found that *PCAT14* expression was able to distinguish tumor from normal (AUC 0.863) (Figure 5*D*) and was high in Gleason-6 with minimal expression noted in benign tissue or Gleason 8 disease (Figure 5*E*).

Functional Evaluation of PCAT14

Since expression of *PCAT14* was lower in high grade prostate cancer and its expression predicted better outcomes, we hypothesize that *PCAT14* may have tumor suppressive effects. To test this hypothesis, we performed overexpression studies in PC3 and LNCaP cells, prostate cancer cell lines that do not express *PCAT14* (Supplementary Figure 2B, C). To overexpress *PCAT14*, we used a CRISPR (clustered regularly interspaced short palindromic repeat)-Cas9 Synergistic Activation Mediator (SAM) complex [18]. This method allows endogenous overexpression of a gene by recruiting artificial transcriptional factors to the promoter using single-guide RNA (sgRNA-MS2) (See method section for details). We designed 6 sgRNAs targeting the *PCAT14* promoter and tested their ability to induce *PCAT14* expression using HEK293 cells stably expressing transcription factors. We found three sgRNAs that significantly increased *PCAT14* expression in HEK293 cells (Supplementary Figure 5A). We next used these sgRNA to construct PC3 and LNCaP cells stable expressing *PCAT14* (Figure 6*A*). Using two independent sgRNAs we were able to achieve 500 to 1000-fold endogenous overexpression of

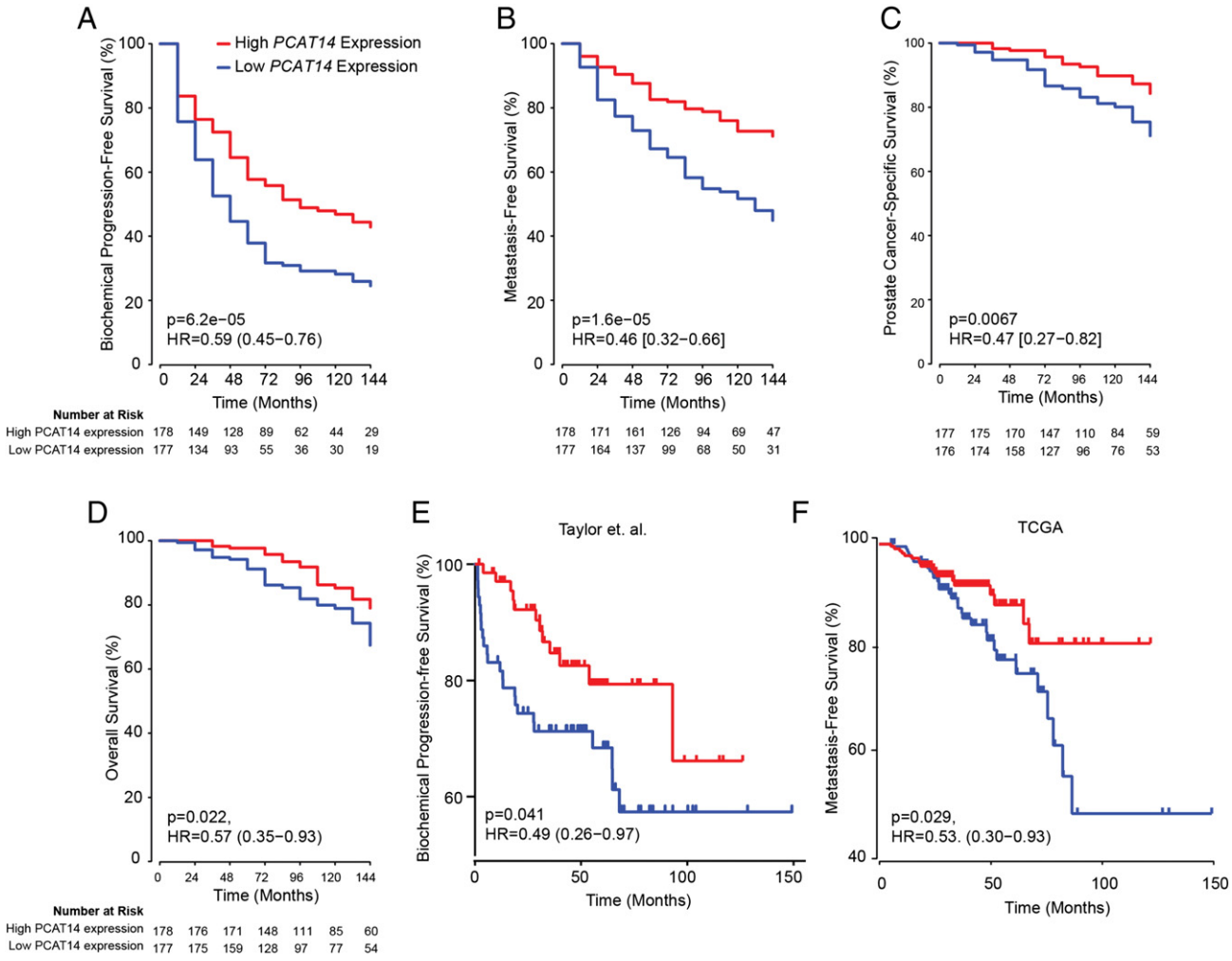


Figure 4. *PCAT14* is a prognostic biomarker. A-D. Kaplan–Meier analyses of prostate cancer outcomes in the John Hopkins cohort. *PCAT14* expression was measured using Affymetrix exon arrays, and subjects were stratified according to their *PCAT14* expression level. Subject outcomes were analyzed for biochemical progression (D) and Metastasis free survival (E), Prostate cancer-specific survival (F) and overall survival (G). Subject outcomes were analyzed for Kaplan–Meier curves, *P* values determined using a log-rank test. E-F. Kaplan–Meier analyses of biochemical progression free survival in the Taylor (E) and Metastasis Free survival in the TCGA (F) cohorts of prostate cancer. Patients were divided into two groups based on the expression level of *PCAT14*. *P* values for Kaplan–Meier curves were determined using a log-rank test.

PCAT14 in PC3 cells (Figure 6*B*) and 20–100 fold overexpression in LNCaP cells (Supplementary Figure 5*B*). While we observed no significant effect of *PCAT14* overexpression on proliferation of PC3 or LNCaP cells (Figure 6*C* and Supplementary Figure 5*C*), overexpression of *PCAT14* lead to suppression of invasion capacity of both PC3 and LNCaP cells (Figure 6*C*, *D*; Supplementary Figure 5*E*, *F*), in line with its prior identified association with clinically indolent disease. We then looked at the effects of *PCAT14* knockdown on cell expressing

Table 1. Multivariate Analysis in JHU Cohort

	Biochemical Recurrence Free Survival		Metastasis Free Survival		Prostate Cancer Free Survival		Overall Survival	
	P-Value	HR [95% CI]	*P*-Value	HR [95% CI]	*P*-Value	HR [95% CI]	*P*-Value	HR [95% CI]
PCAT14 High vs. Low	.00126	0.64 [0.49–0.84]	.000609	0.52 [0.36–0.76]	.0385	0.55 [0.31–0.97]	.0567	0.62 [0.38–1.01]
Age	.818	1 [0.98–1.02]	.65	0.99 [0.96–1.02]	.338	0.98 [0.93–1.02]	.151	0.97 [0.93–1.01]
PSA Int vs. Low	.241	0.83 [0.62–1.13]	.353	0.83 [0.55–1.24]	.385	0.75 [0.4–1.42]	.366	0.77 [0.44–1.35]
PSA High vs. Low	.916	0.98 [0.63–1.52]	.574	0.84 [0.47–1.52]	.463	0.73 [0.31–1.7]	.582	0.81 [0.39–1.7]
Gleason High vs. Low	2.98E-05	1.83 [1.38–2.43]	1.00E-08	3.08 [2.1–4.52]	.000224	3.1 [1.7–5.65]	.000988	2.38 [1.42–3.99]
Seminal vesicle invasion	.0042	1.52 [1.14–2.03]	.453	1.16 [0.79–1.69]	.774	0.92 [0.51–1.66]	.82	0.94 [0.56–1.59]
Surgical margin status	.000533	1.78 [1.28–2.47]	.000276	2.15 [1.42–3.25]	.0487	1.93 [1–3.7]	.0825	1.67 [0.94–2.99]
Extracapsular extension	.456	1.14 [0.81–1.58]	.459	1.21 [0.73–2.03]	.636	0.83 [0.39–1.77]	.816	0.93 [0.48–1.77]
Lymph node invasion	8.98E-12	3.23 [2.31–4.52]	.000164	2.21 [1.46–3.35]	.0616	1.86 [0.97–3.57]	.254	1.42 [0.78–2.6]

HR: Hazard Ratio

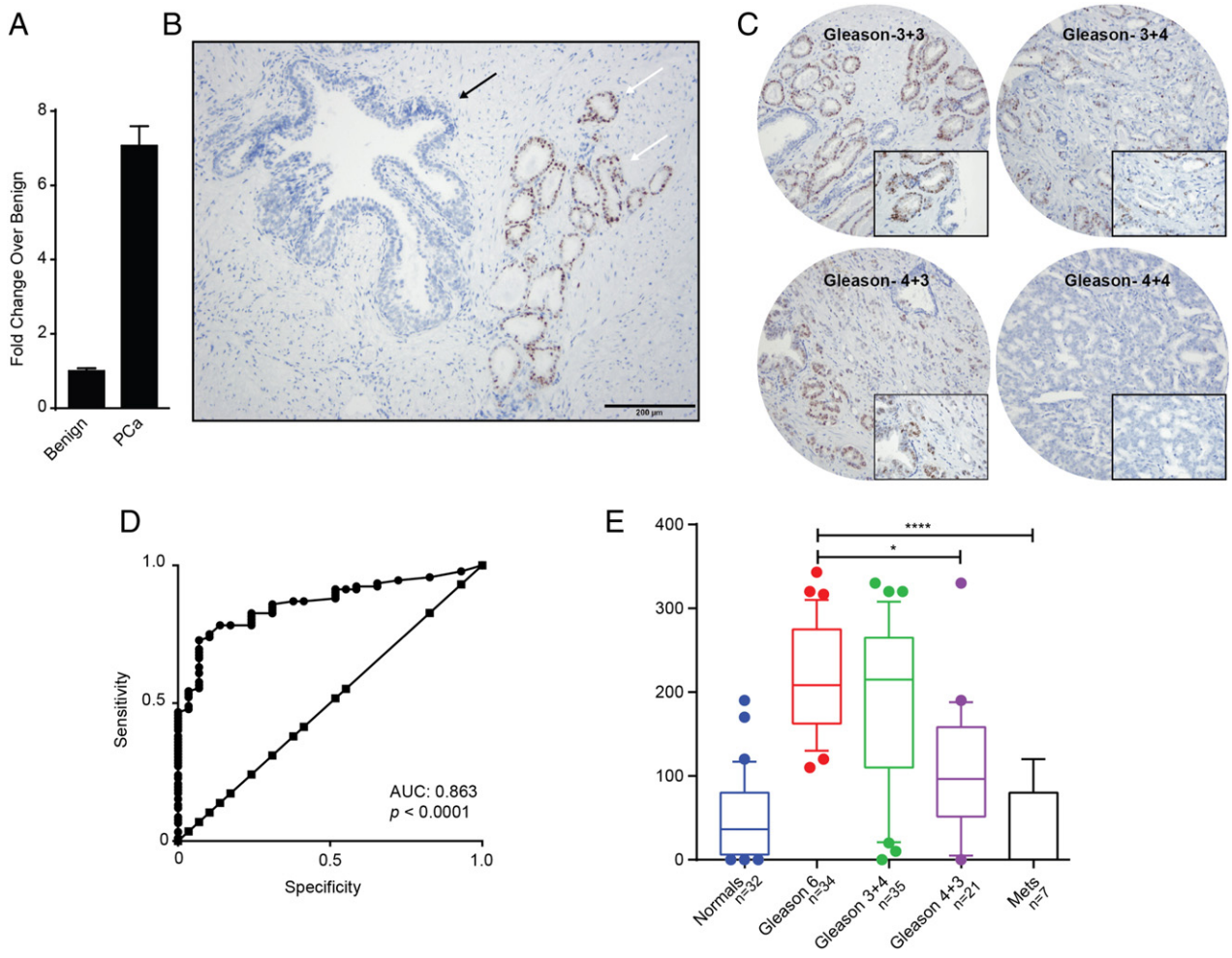


Figure 5. *PCAT14* RNA-ISH in prostate cancer tissues. A. Barplot to show the expression of *PCAT14* in tumor tissue and adjacent benign by qRT-PCR. B. A representative *PCAT14* RNA in-situ hybridization image. White arrows indicate Gleason score 6 disease and black arrows indicate benign glands. C. Representative *PCAT14* In situ hybridization images of human prostate cancer samples of different Gleason grades. D. ROC analysis of *PCAT14* expression in the prostate TMAs. E. Representation of mean *PCAT14* ISH product score for benign prostatic glands (benign), Gleason score 6, Gleason score 3 + 4 = 7, Gleason score 4 + 3 = 7 and Gleason score 8+ clinically localized prostate cancer in a TMA cohort. (** = $P < .01$; compared to benign).

PCAT14 at high levels (VCaP and MDA-PCa-2B). In both MDA-PCa-2b and VCaP cells using 2 independent siRNA as well as 8 independent ASOs we were able to achieve more than 80% knockdown efficiency (Supplementary Figure 5G-J). However, we did not observe a consistent effect on cell proliferation as well as cell invasion (Supplementary Figure 5K-N and data not shown).

Discussion

In this study, we perform a large-scale RNA-sequencing-based analysis of biomarkers associated with indolent versus aggressive prostate cancer and identify the long noncoding RNA *PCAT14* as a marker of low grade and indolent disease. We define the exon structure of *PCAT14* and demonstrate that *PCAT14* is an AR-regulated lncRNA. Using two independent data sets, we show that *PCAT14* is highly upregulated in prostate cancer compared to benign tissue and is able to distinguish prostate cancer from normal tissue with high sensitivity and specificity, suggesting that *PCAT14* can be an excellent diagnostic biomarker. Moreover, we demonstrate that expression of *PCAT14* is prognostic of outcome and is associated with better biochemical progression-free survival, metastases-free survival, and prostate cancer-specific survival. Importantly, we find that *PCAT14* expression is a prognostic biomarker which adds to standard clinicopathologic variables.

As such, *PCAT14* represents a unique biomarker. Most diagnostic biomarkers, such as *PCA3*, can distinguish cancer from normal tissue, but are not prognostic [4]. Conversely, many prognostic biomarkers, such as Ki-67, hold little diagnostic value. It is unclear why *PCAT14* increases significantly in expression during the initial formation of cancer, but then subsequently decreases in expression in disease aggressiveness; this observation requires follow up with further mechanistic studies but is also a feature that gives *PCAT14* value as a biomarker across multiple clinical contexts. Of note, *PCAT14* was also found to be expressed in testicular cancer samples along with prostate cancer, suggesting the role of *PCAT14* in the testicular cancer pathogenesis. However, due to lack of normal testis samples in the TCGA database, it is unclear, at this point, whether *PCAT14* is differentially regulated in testicular cancer compared to normal testis. Recently, the Genotype-Tissue Expression (GTEx) program has

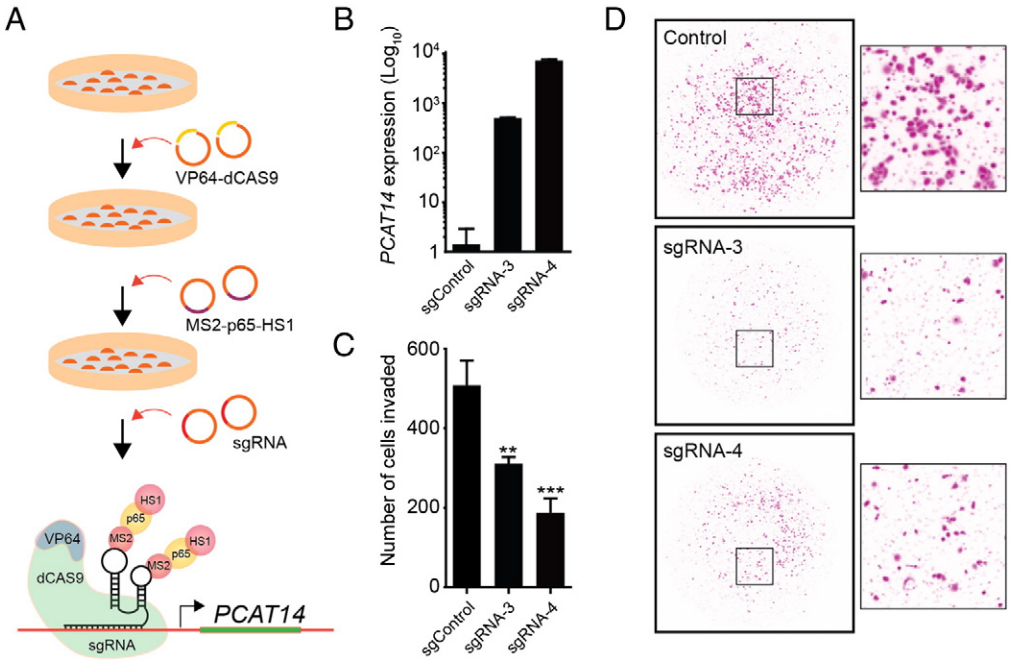


Figure 6. Functional analysis is *PCAT14*. A. Schematic representation of the workflow to endogenously overexpress *PCAT14* in prostate cancer cells using CRISPR/SAM system. B. Bar plots represent fold increase in *PCAT14* level in PC3 cells expressing dCas9–VP64 and MS2–p65–HSF1 with control or 2 independent *PCAT14* sgRNAs. C. Bar plot represent quantification of invaded PC3 cells with or without *PCAT14* expression. D. Representative images of invaded PC3 cells with or without *PCAT14* expression.

generated a large amount of high throughput sequencing data on normal tissue including testis [27]. This data would be useful to look at the role of PCAT14 in testicular carcinoma.

In an attempt to develop a clinical grade assay to detect expression of *PCAT14*, we developed a novel assay, using ISH probes, which can be applied to formalin fixed paraffin-embedded tissues. This ISH assay provides an opportunity to validate our findings in larger cohorts with associated clinical data in the future. Ultimately, an optimized approach for predicting indolent versus aggressive disease will include both clinicopathologic parameters integrated with molecular biomarkers. It is likely that this molecular assay will involve multiplexing multiple biomarkers, and may require combining both tissue-based and urine-based biomarkers. Potential intriguing subsequent studies include the assessment of *PCAT14* and other candidate lncRNAs, in addition to *PCA3*, as urine biomarkers.

There are a several limitations to our study. While we demonstrate the potential value of *PCAT14* expression as a biomarker, it is unclear how *PCAT14* is modulating oncogenic phenotypes, from a mechanistic perspective. Additionally, while we demonstrate the relative specificity of *PCAT14* for both prostate and testicular cancers, the molecular basis underlying this specificity remains to be elucidated. It is known that AR can regulate expression of genes in both prostatic and testicular tissues, but we do not know whether the relative cancer-specificity can be attributed to AR. Clearly, these are important areas for future study.

Overall, our study highlights the need to look at both conventional protein-coding genes and noncoding genes in the search for optimal biomarkers. To our knowledge, there are approximately 20,000 protein coding genes [28], which comprise 2% of the genome. Given our recent study demonstrating that there are close to 60,000 long noncoding RNAs (lncRNAs) [7], many of which are specific to certain cancers, it is clear that these lncRNAs present a relatively underexplored frontier for biomarker development, and that *PCAT14* may represent an initial candidate to be further explored along this frontier.

Conclusion

By performing differential expression analysis between prostate cancer with low vs high Gleason scores, we identified lncRNA *PCAT14* as a prostate cancer- and lineage- specific biomarker of indolent disease. We show that *PCAT14* is an AR-regulated transcript and its overexpression suppresses invasion of prostate cancer cells. Moreover, in multiple independent datasets, *PCAT14* expression associates with favorable outcomes in prostate cancer and adds prognostic value to standard clinicopathologic variables.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2016.07.001.

Acknowledgments

We thank Sethuramasundaram Pitchiaya, Xia Jiang, Fengyun Su and Ingrid Apel for technical assistance; K. Giles for critically looking over the manuscript and the submission of documents; the University of Michigan Viral Vector Core for generating the lentiviral constructs.

References

[1] Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, and Crawford ED, et al (2004). Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* **350**, 2239–2246.

[2] Andriole GL, Crawford ED, Grubb 3rd RL, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, and Reding DJ, et al (2009). Mortality

results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**, 1310–1319.

[3] Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, and Zappa M, et al (2009). Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**, 1320–1328.

[4] Leyten GH, Hessels D, Jannink SA, Smit FP, de Jong H, Cornel EB, de Reijke TM, Vergunst H, Kil P, and Knipscheer BC, et al (2014). Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *Eur Urol* **65**, 534–542.

[5] Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, Williamsen S, Hodge P, Meinke J, and Blase A, et al (2011). Urine TMPRSS2: ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Sci Transl Med* **3**, 94ra72.

[6] Mohler JL, Armstrong AJ, Bahnson RR, D'Amico AV, Davis BJ, Eastham JA, Enke CA, Farrington TA, Higano CS, and Horwitz EM, et al (2016). Prostate Cancer, Version 1.2016. *J Natl Compr Canc Netw* **14**, 19–30.

[7] Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, and Zhao S, et al (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**, 199–208.

[8] Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, and Kominsky HD, et al (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742–749.

[9] Sahu A, Singhal U, and Chinnaiyan AM (2015). Long noncoding RNAs in cancer: from function to translation. *Trends Cancer* **1**, 93–109.

[10] Cancer Genome Atlas Research Network. Electronic address scmo, Cancer Genome Atlas Research N (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025.

[11] Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, and Attard G, et al (2015). Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228.

[12] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

[13] Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

[14] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, and Lander ES, et al (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550.

[15] Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM, Magi-Galluzzi C, Mehra R, Sahu A, and Siddiqui J, et al (2014). RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SChLAP1. *Lancet Oncol* **15**, 1469–1480.

[16] Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK, Siddiqui J, Cao X, Wei J, and Jiang H, et al (2014). A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. *Neoplasia* **16**, 1121–1127.

[17] Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, and Ghadessi M, et al (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet* **45**, 1392–1398.

[18] Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, and Nishimasu H, et al (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588.

[19] Partridge J and Flaherty P (2009). An in vitro FluoroBlok tumor invasion assay; 2009. J Vis Exp.

[20] Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

[21] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, and Chinnaiyan AM (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6.

[22] Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, Eastham JA, Gopalan A, Pienta KJ, and Shen MM, et al (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651.

[23] Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, and Otte AP, et al (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629.

[24] Asangani IA, Dommeti VL, Wang X, Malik R, Cieslik M, Yang R, Escara-Wilke J, Wilder-Romans K, Dhanireddy S, and Engelke C, et al (2014). Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282.

[25] Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, and Reva B, et al (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22.

[26] Mehra R, Udager AM, Ahearn TU, Cao X, Feng FY, Loda M, Petimar JS, Kantoff P, Mucci LA, and Chinnaiyan AM (2015). Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. *Eur Urol.* pii: S0302-2838(15)01211-7. doi: 10.1016/j.eururo.2015.12.003. [Epub ahead of print].

[27] Consortium GT (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.

[28] Consortium EPBirney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, and Dermitzakis ET, et al (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.

[29] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504.

[30] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, and Brenner JC, et al (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243.

[31] Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, and Bergerheim U, et al (2004). Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A* **101**, 811–816.

[32] Vanaja DK, Cheville JC, Iturria SJ, and Young CY (2003). Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res* **63**, 3877–3882.

[33] Arredouani MS, Lu B, Bhasin M, Eljanne M, Yue W, Mosquera JM, Bubley GJ, Li V, Rubin MA, and Libermann TA, et al (2009). Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. *Clin Cancer Res* **15**, 5794–5802.

[34] Best CJ, Gillespie JW, Yi Y, Chandramouli GV, Perlmutter MA, Gathright Y, Erickson HS, Georgevich L, Tangrea MA, and Duray PH, et al (2005). Molecular alterations in primary prostate cancer after androgen ablation therapy. *Clin Cancer Res* **11**, 6823–6834.

[35] Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, and Richie JP, et al (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203–209.

[36] True L, Coleman I, Hawley S, Huang CY, Gifford D, Coleman R, Beer TM, Gelmann E, Datta M, and Mostaghel E, et al (2006). A molecular correlate to the Gleason grading system for prostate adenocarcinoma. *Proc Natl Acad Sci U S A* **103**, 10991–10996.

[37] Setlur SR, Mertz KD, Hoshida Y, Demichelis F, Lupien M, Perner S, Sboner A, Pawitan Y, Andren O, and Johnson LA, et al (2008). Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J Natl Cancer Inst* **100**, 815–825.

[38] Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, and Pienta KJ, et al (2007). Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39**, 41–51.

nature COMMUNICATIONS

ARTICLE

Received 25 May 2016 | Accepted 1 Aug 2016 | Published 26 Sep 2016

DOI: 10.1038/ncomms12791 OPEN

The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression

Yashar S. Niknafs[1,2,*], Sumin Han[3,*], Teng Ma[3,4], Corey Speers[3,5,6], Chao Zhang[3], Kari Wilder-Romans[3], Matthew K. Iyer[1,7], Sethuramasundaram Pitchiaya[1], Rohit Malik[1], Yasuyuki Hosono[1], John R. Prensner[1], Anton Poliakov[1], Udit Singhal[1,8], Lanbo Xiao[1], Steven Kregel[1], Ronald F. Siebenaler[1], Shuang G. Zhao[3], Michael Uhl[9], Alexander Gawronski[10], Daniel F. Hayes[5,6,11], Lori J. Pierce[3,5,6], Xuhong Cao[1,8], Colin Collins[13], Rolf Backofen[9], Cenk S. Sahinalp[10,12,13], James M. Rae[5,6,11], Arul M. Chinnaiyan[1,2,5,6,8,14,15,**] & Felix Y. Feng[1,3,5,6,**,†]

Molecular classification of cancers into subtypes has resulted in an advance in our understanding of tumour biology and treatment response across multiple tumour types. However, to date, cancer profiling has largely focused on protein-coding genes, which comprise <1% of the genome. Here we leverage a compendium of 58,648 long noncoding RNAs (lncRNAs) to subtype 947 breast cancer samples. We show that lncRNA-based profiling categorizes breast tumours by their known molecular subtypes in breast cancer. We identify a cohort of breast cancer-associated and oestrogen-regulated lncRNAs, and investigate the role of the top prioritized oestrogen receptor (ER)-regulated lncRNA, *DSCAM-AS1*. We demonstrate that *DSCAM-AS1* mediates tumour progression and tamoxifen resistance and identify hnRNPL as an interacting protein involved in the mechanism of *DSCAM-AS1* action. By highlighting the role of *DSCAM-AS1* in breast cancer biology and treatment resistance, this study provides insight into the potential clinical implications of lncRNAs in breast cancer.

[1] Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. [2] Department of Cellular and Molecular Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. [3] Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109, USA. [4] Department of Radiation Toxicology and Oncology, Beijing Key Laboratory for Radiobiology (BKLRB), Beijing Institute of Radiation Medicine, Beijing 100850, P. R. China. [5] Breast Oncology Program, University of Michigan, Ann Arbor, Michigan 48109, USA. [6] Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan 48109, USA. [7] Department of Computational Medicine and Bioinformatics, Ann Arbor, Michigan 48109, USA. [8] Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [9] Department of Computer Science and Centre for Biological Signaling Studies (BIOSS), University of Freiburg, Freiburg 79110, Germany. [10] School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6. [11] Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. [12] School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA. [13] Vancouver Prostate Centre, Vancouver, British Columbia, Canada V6H 3Z6. [14] Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. [15] Department of Urology, University of Michigan, Ann Arbor, Michigan 48109, USA. * These authors contributed equally to this work. ** These authors jointly supervised this work. † Present address: Departments of Radiation Oncology, Urology, and Medicine, Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, California 94115, USA. Correspondence and requests for materials should be addressed to A.M.C. (email: arul@umich.edu) or to F.Y.F. (email: felix.feng@ucsf.edu).

Long noncoding RNAs (LncRNAs) have recently been implicated in a variety of biological processes, including carcinogenesis and tumour growth[1–6]. Operating through a myriad of mechanisms[2], lncRNAs have challenged the central dogma of molecular biology as prominent functional RNA molecules. To investigate the role of lncRNAs in breast cancer, we interrogated the expression of lncRNAs across an RNA-sequencing (RNA-seq) breast tissue cohort comprised of 947 breast samples[7,8]. Previously, in a large-scale *ab initio* meta-assembly study from 6,503 RNA-seq libraries, we discovered ~45,000 of unannotated human lncRNAs[7], and this assembly was utilized for the present study. Building on prior work that has begun to investigate the role of lncRNAs in breast cancer[9], we set out to perform a comprehensive analysis of breast cancer tissue RNA-seq data to identify the lncRNAs potentially involved in breast cancer.

Patients with oestrogen receptor (ER)-positive breast cancer have better prognosis than those with ER-negative disease, based on both a more indolent natural history but perhaps more importantly due to effective anti-oestrogen, also designated 'endocrine,' therapy[10]. Despite the efficacy of endocrine therapy, however, the majority of breast cancer deaths occur in women with ER-positive breast cancers, because the incidence of ER-positive versus-negative disease is much higher (approximately 80 versus 20%), and because a substantial fraction of women either have inherent or acquired endocrine therapy-resistant disease[11].

Taken together, these considerations highlight the pressing need to understand the biology of the ER-driven breast cancers and their mechanism of resistance to endocrine therapy. The mechanism through which ER mediates cancer initiation and progression is an area of intense scientific investigation[12–14] that remains incompletely understood. In this regard, while substantial research has been focused on ER abnormalities, such as mutations in the gene encoding for ER (*ESR1*)[14,15] and on the co-existing activation pathways that might mediate resistance, such as HER2[16], few studies exist that interrogate ER-regulated noncoding RNAs[17–21]. Therefore, we set out to perform a comprehensive discovery and investigation of those lncRNAs that are driven by oestrogen in breast cancers drawing from a large human tissue RNA-seq cohort.

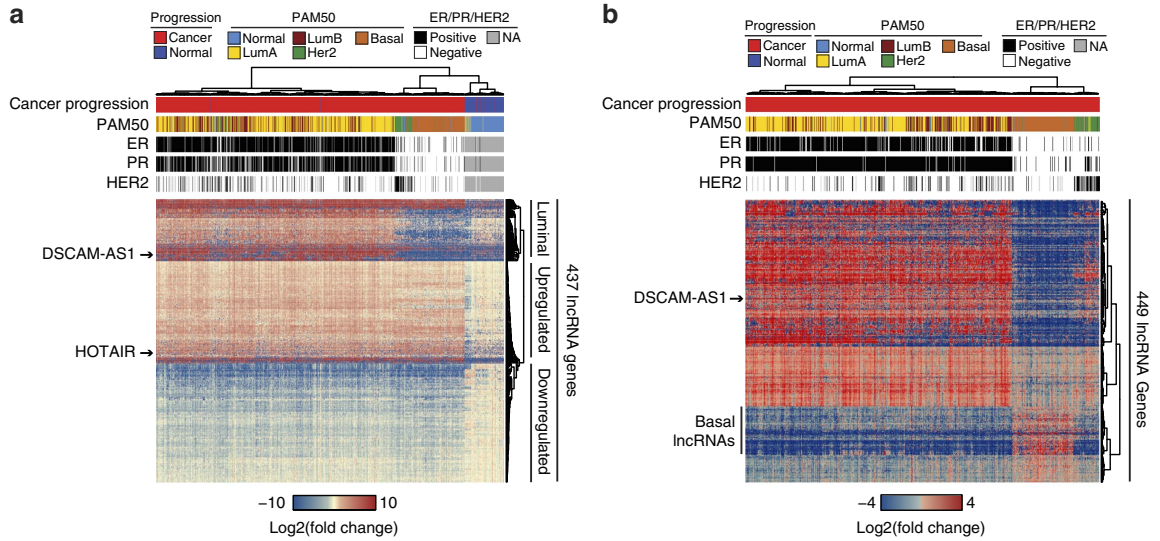
Results

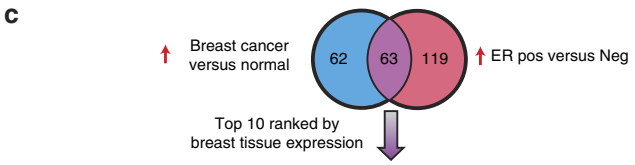
Identification of ER- and breast cancer-associated lncRNAs. We initially focused on those lncRNAs most differentially expressed in breast cancers in comparison to benign adjacent tissue (Supplementary Data 1), utilizing a non-parametric differential expression tool for RNA-seq called Sample Set Enrichment Analysis (SSEA)[7]. After applying an expression filter (at least one fragments per kilobase of transcript per million mapped reads (FPKM) expression in the breast samples in the top 5% based on gene expression level), we identified 437 of the most differentially expressed lncRNAs in breast cancer (Supplementary Data 2). Interestingly, unsupervised hierarchical clustering of the samples based on expression of these lncRNAs across all breast cancer samples (Methods section) largely separated out the breast cancer samples by PAM50 subtypes[22,23], suggesting that lncRNAs may be contributing to the distinct biology of these subtypes (Fig. 1a). While lncRNA expression was unable to distinguish between the ER-driven luminal A and luminal B subtypes, the luminal subtypes were well separated from the HER2, basal and normal subtypes (Fig. 1a). In addition to separating out the clinical subtypes of breast cancer, the lncRNAs themselves separated into three distinct clusters. The first cluster (Fig. 1a, 'Luminal') contains lncRNAs overexpressed mostly in luminal A and luminal B samples, with little expression in samples of the other subtypes, and little expression in normal samples. The next cluster contains lncRNAs upregulated across all breast cancer samples (Fig. 1a, 'Upregulated'), and this cluster included the known breast cancer lncRNA, HOTAIR. The third cluster (Fig. 1a, 'Downregulated') contains lncRNAs downregulated in breast cancers. The lncRNAs in the luminal cluster present a particularly intriguing class of potentially oestrogen-responsive lncRNAs.

Using the 947 breast tumour RNA-seq samples (Supplementary Data 1), we identified lncRNAs differentially expressed in ER-positive versus ER-negative breast tumours (Fig. 1b, Supplementary Data 2). As expected, the expression of lncRNAs differentially expressed in ER-positive tumours separated the luminal tumours from the basal and HER2 on unsupervised hierarchical clustering (Fig. 1b). Quite interestingly, a number of lncRNAs that were downregulated in ER-positive samples exhibited increased expression in the basal samples (Fig. 1b, 'Basal lncRNAs'). While these basal lncRNAs were identified in an ER-positive versus ER-negative cancer analysis, a number of them also exhibit low expression in normal breast tissue (Supplementary Fig. 1). Given that a paucity of known driver genes exist for basal breast cancers and that these tumours are the most clinically aggressive, these basal-specific lncRNAs may represent an exciting future area for basal breast cancer biology.

We set out to investigate potentially oncogenic ER-regulated lncRNAs by intersecting the lncRNAs upregulated in both the cancer versus normal (Fig. 1a) and ER-positive versus ER-negative (Fig. 1b) analyses. Sixty-three lncRNAs were upregulated in both the cancer versus normal analysis and the ER-positive versus ER-negative analysis (Supplementary Data 2, Fig. 1c). To prioritize the most biologically and clinically relevant lncRNAs, we focused on lncRNAs most highly expressed in breast cancer tissues, and those most directly regulated by ER, based on ER binding to the targets' promoter as well as the degree of induction of expression following oestrogen stimulation in breast cancer cells (Fig. 1c and Supplementary Fig. 2a). This approach nominated *DSCAM-AS1* as a lncRNA expressed at a very high level in breast cancer tissues, containing ER promoter binding, and exhibiting the strongest oestrogen induction in MCF7 and T47D cells by both RNA-seq and quantitative PCR (qPCR) validation (Fig. 1c and Supplementary Fig. 2a). We thus selected *DSCAM-AS1* for further investigation.

Characterization of *DSCAM-AS1*. *DSCAM-AS1* has been previously reported to be involved in the proliferation of a luminal breast cancer cell line[20]. It exhibits a highly cancer-specific expression pattern, mostly in breast cancer and lung adenocarcinoma, in transcriptome sequencing data from a cohort of 6,503 cancer and normal tissues and cell lines from the TCGA and the Michigan Center for Translation Pathology[7] (Fig. 2a). Supporting its association with ER biology, *DSCAM-AS1* expression is highly enriched (Student's *t*-test, P value $< 10E^{-5}$) in ER-positive tumours among the breast cancer samples in this RNA-seq cohort with ER status determined by IHC (Fig. 2b and Supplementary Data 1). In addition, analysis of RNA-seq performed on 50 breast cancer cell lines[24] revealed that expression of *DSCAM-AS1* is highly specific to ER-positive cell lines (Fig. 2c and Supplementary Fig. 2b). Further supporting the association of ER with *DSCAM-AS1*, ER chromatin immunoprecipitation-sequencing (ChIP-seq) in both MCF7 and T47D identified ER binding to the *DSCAM-AS1* promoter following oestrogen stimulation (Fig. 2d), and this finding was confirmed by ChIP-qPCR of the *DSCAM-AS1* promoter





MiTranscriptome name	Genomic coordinates	Breast Ca versus norm SSEA %ile	ER+ versus ER- SSEA %ile	Breast expression 95th %ile	ER promoter binding	ER response
DYNLL1-AS1	chr12:120899317-120933832(–)	99.03	99.89	231.53	T47D	↑
CERS6-AS1	chr2:169627826-169678566(–)	99.71	99.93	115.62	⦸	↑
DSCAM-AS1	chr21:41754824-41757280(+)	98.02	97.51	45.29	BOTH	↑↑↑
BRCAT1	chr10:8768317-9084377(+)	99.71	99.53	42.65	BOTH	—
CAT1222	chr10:5500354-5543763(–)	98.71	99.06	35.10	MCF7	↑
BRCAT54	chr5:44742417-44809099(–)	97.97	99.82	30.10	MCF7	—
BRCAT32	chr6:151913411-151937977(–)	98.27	99.86	29.33	⦸	—
CAT146	chr1:197122504-197127797(+)	98.16	97.91	26.80	⦸	↑
CAT659	chr5:44826244-44828803(–)	97.68	98.87	25.55	⦸	—
CAT505	chr4:780400-792957(+)	97.54	99.73	15.71	BOTH	—

Figure 1 | Identification of ER and breast cancer-associated lncRNAs. (**a**) Heatmap depiction of the top cancer versus normal differentially expressed lncRNAs among the TCGA breast RNA-seq cohort ($n=946$). 437 lncRNAs were differentially expressed with an SSEA FDR<1e-5 and an SSEA percentile cutoff of 0.975 (Methods section). Expression values are depicted as log2 of the fold change over the median of the normal samples ($n=104$). Unsupervised hierarchical clustering was done on both lncRNAs and patients. Cancer progression, PAM50 classification, and ER, PR, and HER2 status are shown above heatmap. LncRNAs clustered into 3 distinct categories, 'Luminal', 'Upregulated', and 'Downregulated'. Two representative lncRNAs are highlighted. (**b**) Heatmap depiction of the top ER-positive versus ER-negative lncRNAs. 449 lncRNAs met the SSEA criteria described in **a**. Unsupervised clustering was performed for samples and lncRNAs. Expression values depicted as log2 of the fold change over the median of the ER-negative samples ($n=538$). Cancer progression, PAM50 classification, and ER, PR and HER2 status are shown above heatmap. One representative lncRNA is highlighted along with a group of lncRNAs with basal-specific expression. (**c**) Venn diagram of the intersection of the breast cancer versus normal and ER-positive versus ER-negative analyses. Intersection is shown for the overexpressed lncRNAs in both categories. The top 10 lncRNAs based on expression level in breast cancer tissues (expression value of 95th percentile sample) are depicted in table. ER promoter binding determined via ChIP-seq is depicted (in either MCF7, T47D cell lines, or both) along with expression response from RNA-seq following 3 h of oestrogen stimulation in MCF7 cells (one arrow represents >1.5 fold increase, three arrows represents >2.5 fold increase).

(Supplementary Fig. 2c). The isoforms of *DSCAM-AS1* in MCF7 cells were identified using 3′ and 5′ RACE (Fig. 2d and Supplementary Table 1). *DSCAM-AS1* expression is induced in both MCF7 and T47D cells after oestrogen stimulation, and this induction is reversed with addition of tamoxifen, corroborating that ER is in fact regulating the expression of this lncRNA (Fig. 2e). Expression of known ER-regulated protein-coding genes *GREB1* and *PGR* follow the same pattern of response to oestrogen, while the lncRNA *MALAT1*, serving as a negative control, is not induced by oestrogen (Fig. 2e). In addition to being oestrogen-responsive, *DSCAM-AS1* expression is present in both the cytoplasm and nucleus at nearly identical fractions in both MCF7 and T47D cells (Supplementary Fig. 2d), and the identity of *DSCAM-AS1* as a noncoding gene was corroborated using the CPAT tool[25] (Supplementary Fig. 2e). We used single-molecule fluorescence *in situ* hybridization (ISH) to further dissect the subcellular localization and gene expression levels of DSCSM-AS1 in breast cancer cells. To this end, we designed probes that targeted all potential isoforms of the transcript predicted by RACE. On staining, we found that each MCF7 cell expressed ~800 copies of the DSCAM-AS1 transcript, almost half as much as the expression level of GAPDH (Supplementary Fig. 2f,g), additionally the similar nuclear and cytoplasmic localization was corroborated by ISH (Supplementary Fig. 2h). While the abundance of DSCAM-AS1 was lower in T47D cells (~260 molecules per cell, Supplementary Fig. 2i,j), the relative expression level (compared with GAPDH) and the subcellular localization pattern were very

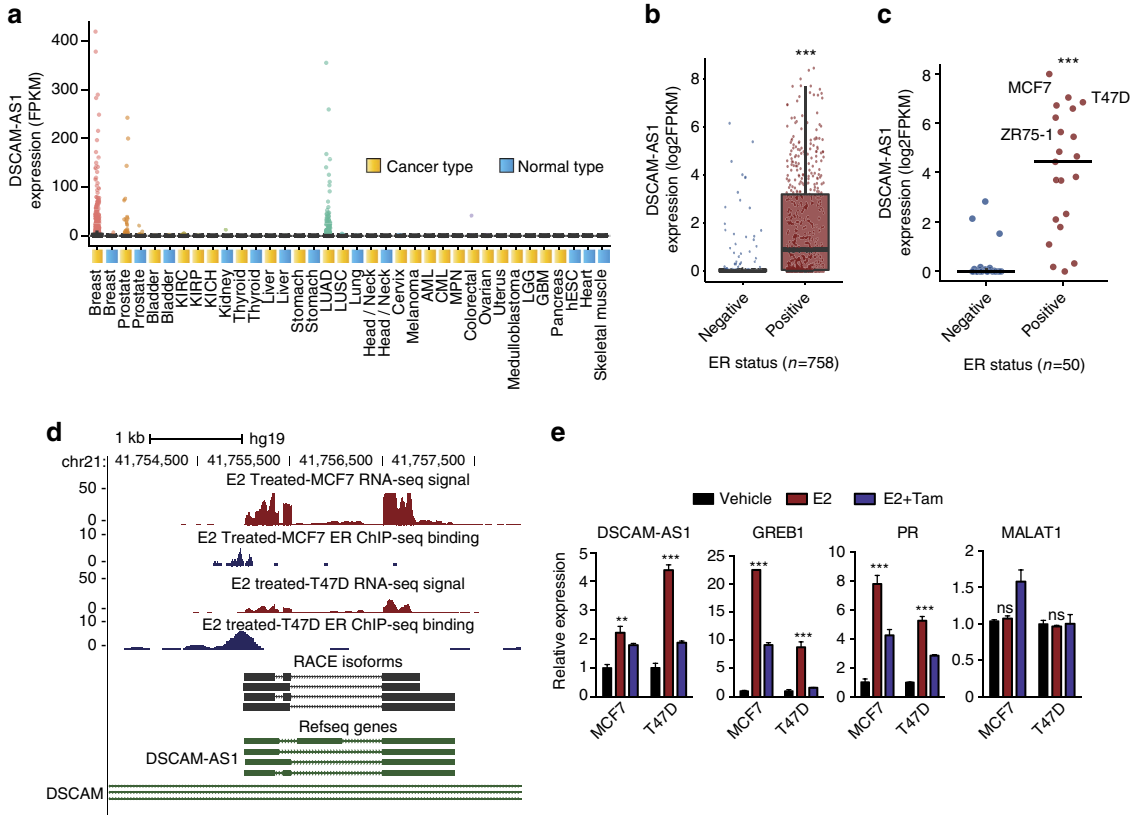


Figure 2 | Characterization of *DSCAM-AS1*. (**a**) Plot highlighting the expression in FPKM of *DSCAM-AS1* in the 6,503 sample MiTranscriptome RNA-seq compendium[7] categorized by the different cancer/tissue types. Each point represents one RNA-seq tissue sample. (**b**) Expression of *DSCAM-AS1* is significantly higher in ER-positive breast cancer tissue samples ($n=584$) compared with ER-negative samples ($n=174$). Expression was analysed in samples for which ER IHC was performed. Each point represents one RNA-seq sample. ***$P<0.0001$, comparing ER-positive with -negative. (**c**) Expression of *DSCAM-AS1* by RNA-seq in breast cancer cell lines categorized by ER status. *DSCAM-AS1* expression is significantly higher in ER-positive cell lines ($n=21$) versus ER-negative cell lines ($n=29$). Each point represents one cell line. ***$P<0.0001$, comparing ER-positive to –negative via Student's *t*-test. (**d**) UCSC genome browser depiction of *DSCAM-AS1* region on chromosome 21. RNA-seq expression track shown in red, and ER ChIP-seq shown in blue. Refseq transcripts shown in green. RACE verified transcript structure shown in black. (**e**) qPCR expression of *DSCAM-AS1*, *GREB1*, *PGR*, and *MALAT1* 8 h following addition of DMSO vehicle (black), 10 nM estrogen (red), and 10 nM estrogen and 1 μM tamoxifen (blue) in MCF7 and T47D cell lines. Error bars represent s.e.m. for three biological replicates. **$P<0.001$, ***$P<0.0001$, NS: $P>0.01$ comparing with vehicle for each condition via Student's *t*-test. NS, not significant.

similar to those observed in MCF7 cells (Supplementary Fig. 2k).

***DSCAM-AS1* is implicated in cancer aggression**. We next investigated the clinical relevance of *DSCAM-AS1*. Given that *DSCAM-AS1* is a lncRNA, its expression is not measured by most traditionally used microarrays, which are the primary high-throughput platforms annotated with reliable clinical outcomes in breast cancer[26]. As a surrogate, we employed a guilt-by-association analysis to interrogate the clinical relevance of those genes most correlated to *DSCAM-AS1*. Given that *DSCAM-AS1* is an ER-regulated lncRNA, correlation was performed using only ER-positive breast cancers, to ascertain clinical relevance in the breast cancer samples in which *DSCAM-AS1* would be enriched and most relevant. We obtained a number of breast cancer clinical data sets from Oncomine[27] containing gene expression sets associated with the presence of cancer (versus normal tissue), high clinical grade, recurrence, survival and metastasis[22,23,26–40] (Methods section). We assessed for the overlap between these gene sets with the genes most positively or negatively correlated to *DSCAM-AS1*. *DSCAM-AS1* positively correlated genes were significantly associated with clinical signatures associated with increased cancer aggression, tamoxifen resistance, higher grade, stage and metastasis (Fig. 3a,b, Supplementary Data 3 and 4). Similarly, the *DSCAM-AS1* negatively correlated genes associated with clinical signatures that portended a more favourable clinical outcome (Supplementary Fig. 3a,b, Supplementary Data 3 and 4). For many of the clinical concepts, *DSCAM-AS1* positively correlated genes displayed a clinical association comparable to those genes most correlated to *EZH2*, a gene known to be a marker of clinical aggressiveness in breast cancer[41], while genes correlated to other lncRNAs expressed in breast tissue, such as *HOTAIR*, *MALAT1* and *NEAT1*, showed modest-to-no association (Fig. 3b, Supplementary Fig. 3b, Supplementary Data 3 and 4). In addition, performing a Gene Set Enrichment Analysis (GSEA)[42] on all genes correlated to *DSCAM-AS1* yielded significant association with a myriad of breast cancer, cancer aggressiveness, and ER- and tamoxifen-associated gene signatures (Supplementary Fig. 3c). While ER-positive breast cancers typically result in better clinical outcomes[23], among the luminal breast cancers, *DSCAM-AS1* is expressed significantly higher in luminal B, a clinical subtype containing most of the clinically aggressive ER-positive breast cancers[22,23] (Fig. 3c). Despite these associations of clinical aggression with *DSCAM-AS1*, in a survival analysis of the ER-positive TCGA breast samples, expression of *DSCAM-AS1* was not significantly associated with clinical outcome (Supplementary Fig. 3d). Definitive assessment of

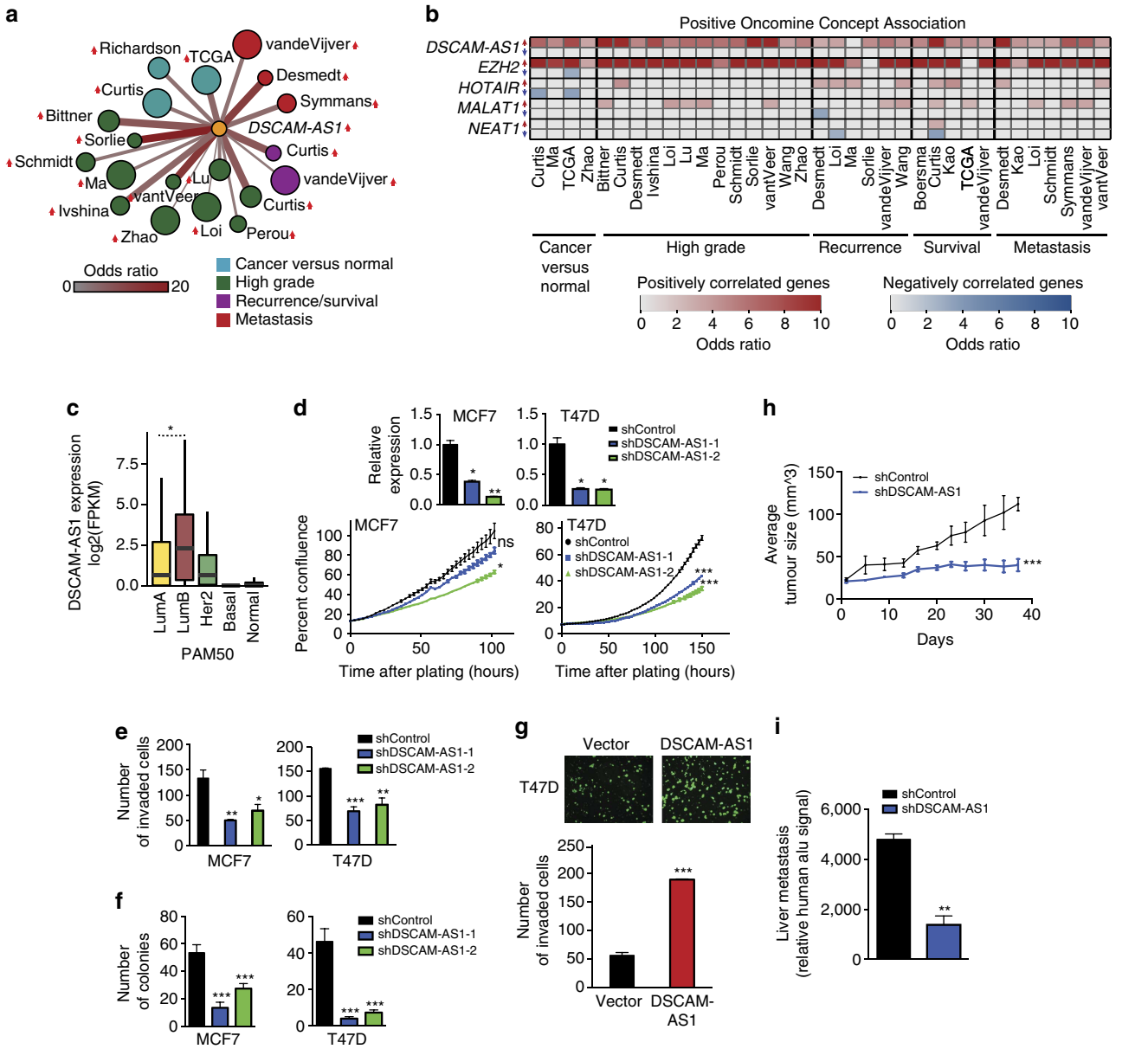


Figure 3 | *DSCAM-AS1* is implicated in cancer aggression clinically and in cell lines. (**a**) Cytoscape depiction of the overlap between the 150 genes most positively correlated with *DSCAM-AS1* and clinical signatures from Oncomine[27] for breast cancer clinical outcomes (i.e., recurrence, survival and metastasis), high cancer grade, and cancer versus normal. All significant associations with an odds ratio >6 are shown (Fisher's *P* value $<1E^{-4}$). Size of node reflects the size of the gene signature, and the thickness/redness of the line represents the magnitude of the odds ratio. (**b**) Heatmap displaying the overlap between the top 150 genes correlated to *DSCAM-AS1*, *EZH2*, *HOTAIR*, *MALAT1* and *NEAT1* and the genes positively associated with various breast cancer clinical signatures (see above). For each gene, the top row depicts the odds ratio for the positively correlated genes (red), and the bottom row represents the odds ratio for the negatively correlated genes (blue). The first name of the author for each clinical study is listed. (**c**) Expression of *DSCAM-AS1* from breast cancer RNA-seq by PAM50 classification ($n=946$). Luminal B expression is significantly greater than Luminal A (Student's *t*-test, $P=0.006$) (**d**) Incucyte proliferation assay performed following knockdown of *DSCAM-AS1* using two independent shRNAs. Degree of knockdown determined by qPCR shown above. Error bars represent the s.e.m. for three biological replicates. $^{*}P<0.01$, $^{**}P<0.001$, $^{***}P<0.0001$, NS: $P>0.01$ comparing to shControl for each condition via Student's *t*-test. (**e**) Invasion assay following shRNA knockdown of *DSCAM-AS1* using Matrigel coated Boyden chamber assay. Error bars represent the s.e.m. for three biological replicates. $^{*}P<0.01$, $^{**}P<0.001$, $^{***}P<0.0001$ comparing to shControl for each condition via Student's *t*-test. (**f**) Soft agar colony formation assay following shRNA knockdown of *DSCAM-AS1*. Error bars represent the s.e.m. for three biological replicates. $^{***}P<0.0001$ comparing to shControl for each condition via Student's *t*-test. (**g**) Invasion assay following overexpression of *DSCAM-AS1* and *LacZ* control. Error bars represent the s.e.m. for three biological replicates. Representative invasion images shown above. $^{***}P<0.0001$, comparing to vector overexpression via Student's *t*-test. (**h**) Mouse xenograft study of tumour growth for T47D cells with shRNA knockdown of *DSCAM-AS1*. Xenografts with shRNA knockdown of *DSCAM-AS1* ($n=24$) exhibited reduced growth when compared to control shRNA knockdown ($n=26$). Error bars represent the s.e.m. for all xenografts used. $^{***}P<0.0001$, comparing to shControl via Student's *t*-test. (**i**) Assessment of xenograft metastasis to liver by human Alu PCR, which detects the human cancer cells in the mouse liver. Error bars represent the s.e.m. for three biological replicates. $^{**}P<0.001$ comparing with shControl via Student's *t*-test. NS, not significant.

survival in this cohort, however, will likely require more robust and longer-term clinical curation of the TCGA breast samples.

We then studied the role of *DSCAM-AS1* on oncogenic phenotypes in ER-positive breast cancer cell lines. In MCF7 and T47D cells, stable knockdown of *DSCAM-AS1* was achieved using shRNA approaches. *DSCAM-AS1* knockdown reduced the proliferative ability of both cell lines (Fig. 3d), diminished the ability of these cells to invade in a Boyden chamber invasion assay (Fig. 3e), and substantially abolished the ability of these cells to form colonies in soft agar (Fig. 3f). While ER regulates levels of *DSCAM-AS1*, *ER* expression and protein levels are not dependent on level of *DSCAM-AS1* (Supplementary Fig. 4a), ruling out the possibility that the phenotype observed could be explained through changes in the level of *ER*. In addition, knockdown of *DSCAM-AS1* exhibited no affect on RNA or protein levels of the *DSCAM* gene, in which *DSCAM-AS1* resides antisense and intronic (Supplementary Fig. 4b). To further demonstrate the impact of *DSCAM-AS1* on aggressive cancer phenotypes, we overexpressed *DSCAM-AS1* in T47D (Supplementary Fig. 4c) and ZR75-1 (Supplementary Fig. 4d), two ER-positive breast cancer cell lines with moderate *DSCAM-AS1* expression (Fig. 2c), and observed an increase in the invasion phenotype (Fig. 3g and Supplementary Fig. 4e). MCF7 cells were not included in the overexpression studies as *DSCAM-AS1* is already expressed at a very high level in these cells (Fig. 2c). Overexpression was also tested in MDA-MB-231 cells (Supplementary Fig. 4f), a common ER-negative cell line. However, exogenous *DSCAM-AS1* was unable to confer oncogenicity via proliferation (Supplementary Fig. 4g) and invasion (Supplementary Fig. 4h). This phenomenon may be explained by a requisite genetic and epigenetic milieu provided by ER-positive cells in order for *DSCAM-AS1* to confer its cancer phenotype, and more investigation into the precise mechanisms through which it acts will shed light on this finding. Furthermore, the simple presence of *DSCAM-AS1* alone is not sufficient to make cells highly aggressive, as evidenced by its high expression in ER-positive cell lines that are moderately invasive (for example, MCF7). To further characterize the impact of *DSCAM-AS1* on cancer phenotype, we performed a mouse xenograft tumour growth assay, showing that loss of *DSCAM-AS1* reduces the growth of implanted T47D cells *in vivo* (Fig. 3h). The metastatic potential of these implanted cells were also reduced with *DSCAM-AS1* knockdown, as evidenced through decreased liver metastasis following xenograft (Fig. 3i).

Role of hnRNPL in *DSCAM-AS1* mechanism. LncRNAs have been shown to be functional through their binding interactions with other RNAs, DNA, and with proteins[2]. Thus, identifying protein binding partners for *DSCAM-AS1* is a crucial step in determining the mechanism through which it confers oncogenicity. To identify *DSCAM-AS1* binding partners, we performed pull-down of *DSCAM-AS1* and performed mass spectrometry on the pull-down product to identify proteins bound to *DSCAM-AS1* (Fig. 4a). The protein hnRNPL was observed to have the highest spectral counts for the sense form of *DSCAM-AS1* with zero spectral counts in the antisense pull-down (Fig. 4b). In addition, PCBP2, a protein known to complex with hnRNPL[43], was also among the top proteins bound to *DSCAM-AS1*. We thus investigated the interaction between *DSCAM-AS1* and hnRNPL further. HnRNPL is a protein widely expressed in many tissue types (Supplementary Fig. 5a) and has been implicated in regulating RNA stability and processing with subsequent effects on gene expression[44–47]. The binding of hnRNPL to *DSCAM-AS1* was confirmed by RNA pull-down followed by western blot, with no binding of hnRNPL to the negative control antisense transcript (Fig. 4c). Other RNA-binding proteins did not bind *DSCAM-AS1*, however, suggesting that *DSCAM-AS1* does not promiscuously bind to RNA-binding proteins in general (Fig. 4c). To further confirm this binding interaction and its specificity, RNA immunoprecipitation (RIP) was performed with using antibodies directed against hnRNPL. *DSCAM-AS1* was highly enriched by anti-hnRNPL RIP in both MCF7 and T47D cells, while control coding and noncoding genes exhibited modest binding (Fig. 4d). In addition, anti-snRNP70 and anti-HuR RIP failed to pull-down *DSCAM-AS1*, further suggesting the specificity of the *DSCAM-AS1*-hnRNPL interaction (Supplementary Fig. 5b).

To more specifically investigate the functional relationship of *DSCAM-AS1* and hnRNPL, we performed rescue studies assessing the impact of *hnRNPL* knockdown on the invasive advantage conferred by *DSCAM-AS1* overexpression, observing that reduction of *hnRNPL* levels entirely reversed the increase in invasion observed on *DSCAM-AS1* overexpression (Fig. 4e, Supplementary Fig. 6a). Because there was only slight, non-significant reduction in invasion with *hnRNPL* knockdown in control cells, the marked reduction in invasion observed in the *DSCAM-AS1* overexpressing cells with *hnRNPL* knockdown may be the result of *hnRNPL* affecting invasion in a mechanism exclusive to *DSCAM-AS1*. So, to further characterize the functional relationship between *DSCAM-AS1* and hnRNPL, we set out to localize the binding site of hnRNPL within the *DSCAM-AS1* lncRNA. Using *in silico* prediction drawing from prior studies of hnRNPL crosslinking-immunoprecipitation sequencing (CLIP-seq)[48], a single strong predicted binding peak was identified near the 3′-end of *DSCAM-AS1* (Fig. 4f). HnRNPL has been shown to bind CACA-rich RNA sites[45], and the predicted binding region possessed a 10 base pair CACA stretch. To identify if this predicted region does in fact account for the hnRNPL binding, multiple mutant forms of *DSCAM-AS1* were created with or without the binding site. *DSCAM-AS1-5* and *DSCAM-AS1-3* are large deletion mutants containing only the 5′- and 3′-end, respectively, with only *DSCAM-AS1-3* possessing the predicted binding site, and *DSCAM-AS1-D* is a mutant form with the 27 nucleotides comprising the predicted binding site deleted (Fig. 4f,g, red). The various mutant forms of *DSCAM-AS1* were expressed in HEK293, a cell line that lacks endogenous *DSCAM-AS1* expression while still expressing hnRNPL (Supplementary Fig. 6b). While both the full-length and *DSCAM-AS1-3* mutant retained hnRNPL binding, loss of the predicted binding region was effective in abrogating hnRNPL binding via both Western blot following RNA pull-down (Fig. 4g) and by qPCR following hnRNPL RIP (Supplementary Fig. 6c). All deletion mutants were expressed at comparable levels, ruling out the possibility of falsely diminished binding due to failed expression of the mutant construct (Supplementary Fig. 6d). RNA secondary structure is a crucial component of RNA functionality and is a key player in RNA-protein interactions. While the 27 nucleotide deletion in the *DSCAM-AS1-D* mutant is a small fraction of the total number of bases in the transcript, to ensure that this deletion was not causing a marked RNA secondary structure change, we investigated the impact of this deletion on RNA secondary structure via the RNAfold structure prediction tool[49]. Evidenced by a minimal free energy prediction, the posited secondary structure of *DSCAM-AS1* is largely similar to that of *DSCAM-AS1-D* (Supplementary Fig. 6e), suggesting that the loss of hnRNPL binding observed with the *DSCAM-AS1-D* mutant is not due to a dramatic secondary structure rearrangement. Quite interestingly, overexpression of the *DSCAM-AS1-D* mutant in T47D cells failed to recapitulate the increased invasion observed when overexpressing full-length *DSCAM-AS1* (Fig. 4h). This finding, in combination with the rescue studies following *hnRNPL* knockdown (Fig. 4e), strongly

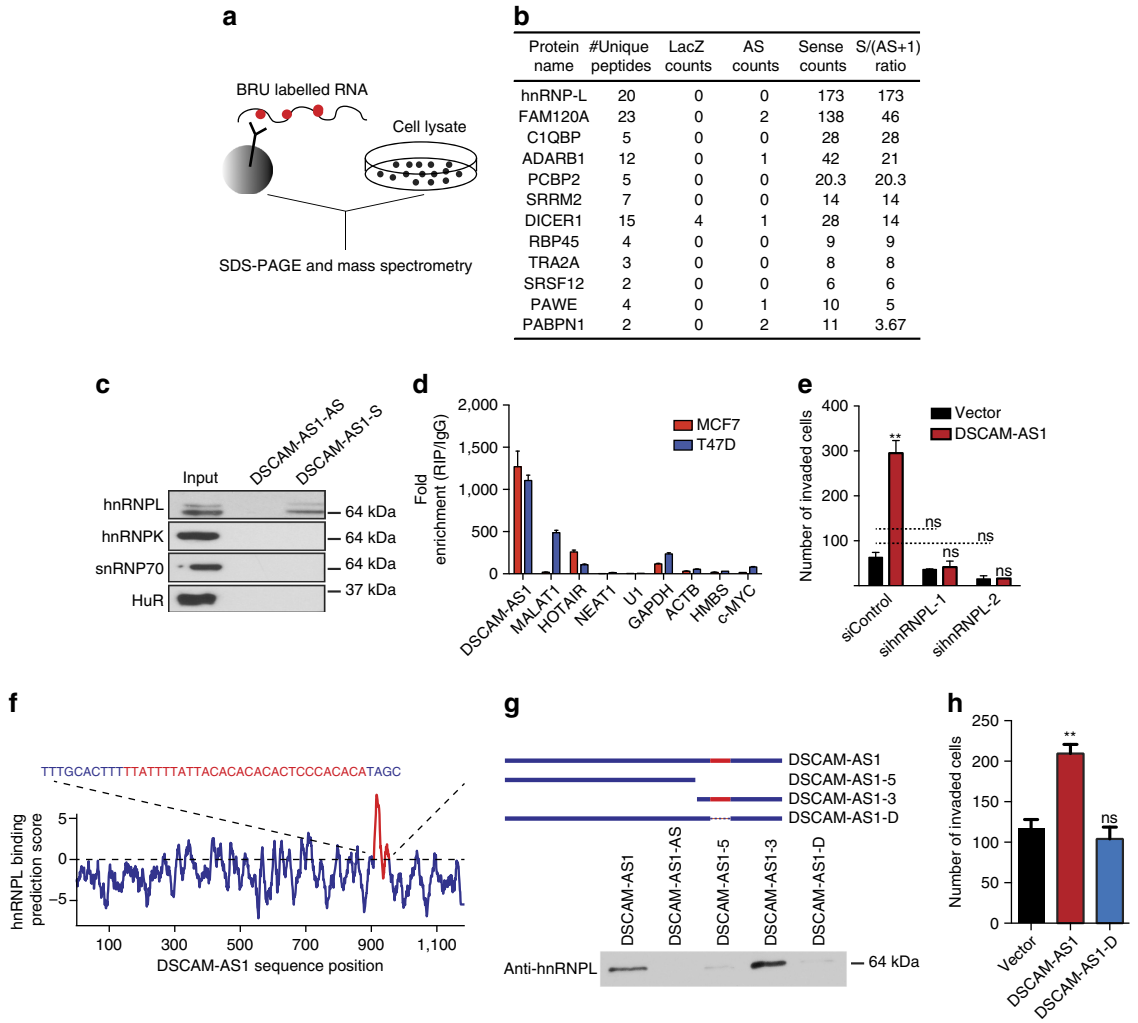


Figure 4 | Physical and functional relationship of *DSCAM-AS1* with hnRNPL. (**a**) Schematic representation of the RNA pull-down technique used to identify protein binding partners of *DSCAM-AS1*. The BRU labeled RNA transcripts are incubated with cell lysate from T47D cells and the eluted protein is resolved by SDS–PAGE. RNA-bound protein product is then processed by mass spectrometry. (**b**) Top protein binding partners for *DSCAM-AS1*. Pull-down of *LacZ* and antisense *DSCAM-AS1* used as control. S/AS ratio determined as sense counts divided by 1+antisense counts. (**c**) Western blot of hnRNPL, hnRNPK, snRNP70, and HuR following pull-down of BRU labelled *DSCAM-AS1* and antisense *DSCAM-AS1*. (**d**) qPCR following RIP for hnRNPL performed in MCF and T47D cells. Data represented as fold-enrichment over IgG RIP. Error bars represent the s.e.m. for three biological replicates. (**e**) Invasion assay for T47D cells overexpressing *LacZ* control or *DSCAM-AS1* following siRNA-mediated knockdown of *hnRNPL*. Error bars represent the s.e.m. for two biological replicates. $^{**}P<0.001$, NS: $P>0.01$ comparing to siControl (unless otherwise specified with dotted line) for each condition via Student's *t*-test. (**f**) Per base *in silico* prediction for binding of hnRNPL to *DSCAM-AS1*. One strong predicted binding peak exists in the 3′ region of *DSCAM-AS1* shown in red. (**g**) Schematic depicting the mutant forms of *DSCAM-AS1* generated with or without the predicted binding site (top). Western blot for hnRNPL shown following pull-down of each mutant form of *DSCAM-AS1* in HEK293 cells (bottom). (**h**) Invasion assay in T47D cells overexpressing *LacZ* control, full-length *DSCAM-AS1*, and the *DSCAM-AS1-D* mutant. Error bars represent the s.e.m. for three biological replicates. $^{**}P<0.001$, NS: $P>0.01$ comparing to vector overexpression for each condition via Student's *t*-test. NS, not significant; siRNA, small interfering RNA.

suggest that *DSCAM-AS1* promotes oncogenicity via its interaction with hnRNPL in these ER-positive breast cancer cells.

Role of *DSCAM-AS1* in tamoxifen resistance. A substantial number of patients with ER-positive breast cancer eventually develop resistance to endocrine therapy and present with clinical recurrence and metastasis[11,50,51]. Thus, as *DSCAM-AS1* is implicated in poor-prognosis ER-positive breast cancer (Fig. 3a-c and Supplementary Fig. 3), we set out to investigate its potential role in subverting oestrogen dependence and promoting resistance to anti-oestrogen therapies. We continuously passaged MCF7 cells in 1 uM tamoxifen for 6 months until we attained a subpopulation of MCF7 cells that were able to grow in in tamoxifen and termed these tamoxifen-resistant MCF7 cells (TamR-MCF7). Interestingly, although expression of canonical ER targets (*GREB1* and *PGR*) was decreased compared to the parental MCF7 cells, *DSCAM-AS1* expression was significantly upregulated despite already being expressed at very high levels in MCF7 cells (Fig. 5a). The levels of *ER* were also increased, which is likely a compensatory upregulation in response to the continual anti-oestrogen effects of tamoxifen. Additionally, short-term tamoxifen treatment of parental MCF7 cells transiently reduced *DSCAM-AS1* levels at 8hrs following tamoxifen treatment, with a rise back to pre-treatment levels after 24 h (Supplementary Fig. 7a). In contrast, canonical ER target, *GREB1*, exhibited pronounced expression reduction at both the short- and long-term timescale (Supplementary Fig. 7b). To interrogate whether

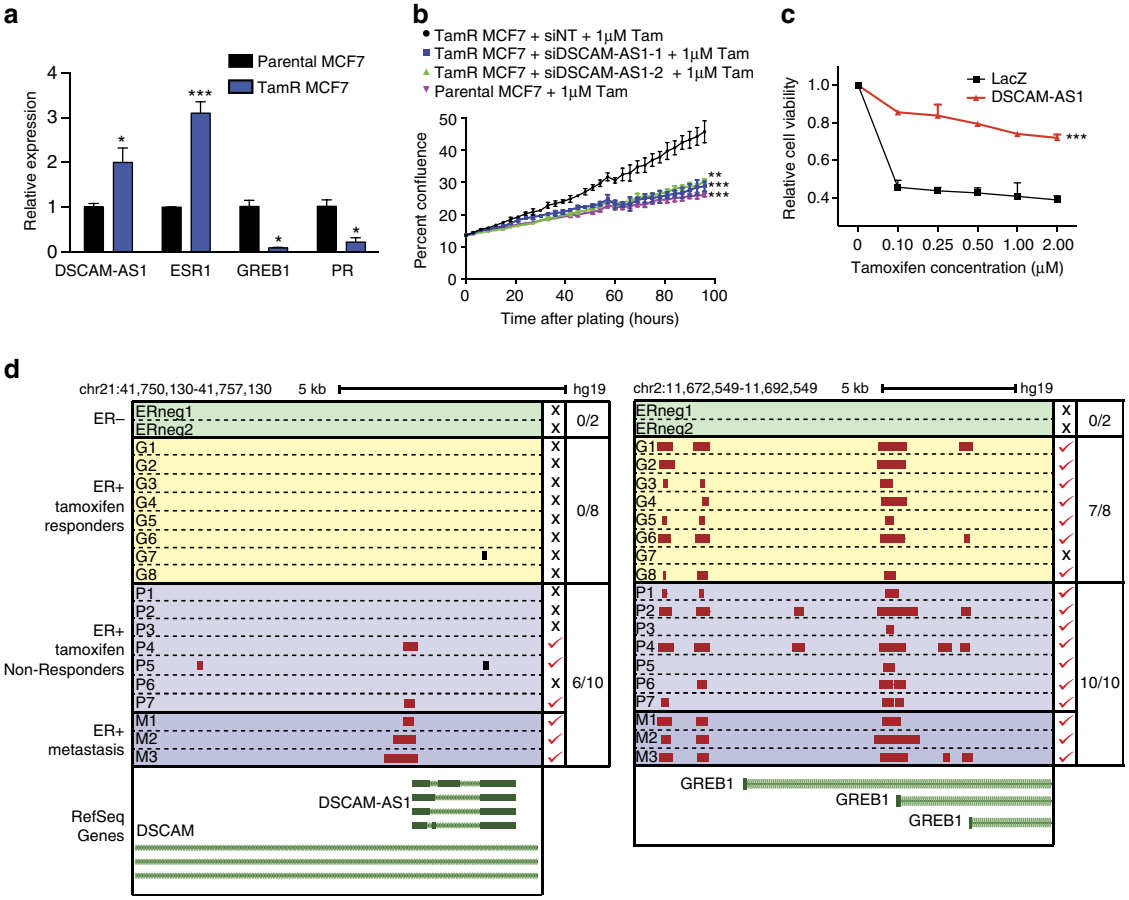


Figure 5 | *DSCAM-AS1* is implicated in tamoxifen resistance. (**a**) qPCR expression of *DSCAM-AS1*, *ESR1*, *GREB1* and *PGR* in tamoxifen-resistant MCF7 cells relative to parental MCF7. Error bars represent the s.e.m. for three biological replicates. $*P<0.01$, $***P<0.0001$, comparing to parental MCF7 for each condition via Student's *t*-test. (**b**) Proliferation assay in parental MCF7 cells and in TamR-MCF7 cells following siRNA-mediated knockdown of *DSCAM-AS1* via two independent siRNAs. Error bars represent the s.e.m. for three biological replicates. $**P<0.001$, $***P<0.0001$, comparing to parental TamR siNT for each condition via Student's *t*-test. (**c**) WST viability assay following 10 days of culture in varying levels of tamoxifen performed for T47D cells overexpressing LacZ control and *DSCAM-AS1*. Error bars represent the s.e.m. for three biological replicates. $***P<0.0001$, comparing to LacZ overexpression via Student's *t*-test. (**d**) Depiction of oestrogen receptor binding to the *DSCAM-AS1* (left) and *GREB1* (right) promoters via ChIP-seq performed in primary and metastatic breast cancer tumour tissues. ER status and response to tamoxifen treatment detailed to left. ER binding peaks (determined using MACS software) are depicted in red (for promoter binding) and black (for non-promoter binding). Promoter defined as 5KB upstream of any transcriptional start site. ER promoter binding indicated by red check or 'x' to the right. Genomic coordinates in hg19 listed above. siRNA, small interfering RNA.

this upregulation of *DSCAM-AS1* in the TamR-MCF7 cells is functionally significant, we assessed the proliferative capacity of these cells following *DSCAM-AS1* knockdown. With knockdown levels of *DSCAM-AS1* comparable to the endogenous levels in parental MCF7 cells (Supplementary Fig. 7c), knockdown of *DSCAM-AS1* in TamR-MCF7 cells led to a loss of their baseline proliferative advantage when cultured in tamoxifen, exhibiting a proliferation profile nearly identical to that of the parental MCF7 cells (Fig. 5b). Additionally, knockdown of *hnRNPL* in these cells produced a similar loss of proliferative capacity in the TamR-MCF7 cells (Supplementary Fig. 7d,e), suggesting that both *DSCAM-AS1* and hnRNPL may be playing a role in promotion of the tamoxifen resistance developed by these cells.

We also interrogated the ability of *DSCAM-AS1* to confer tamoxifen resistance in native T47D cells via overexpression of *DSCAM-AS1*. *DSCAM-AS1* overexpression was also associated with tamoxifen-resistant growth in a dose-dependent manner, with a striking increase in cell viability at levels of tamoxifen as low was 100 nM (Fig. 5c). Additionally, in line with the ability of *DSCAM-AS1* to provide oestrogen-independent growth advantage, cells overexpressing *DSCAM-AS1* also exhibited a proliferative advantage when grown in oestrogen-deprived medium compared to normal serum (Supplementary Fig. 7f). This growth advantage was abolished with the addition of oestrogen, and returned with the subsequent addition of tamoxifen (Supplementary Fig. 7f). Conversely corroborating the relationship of *DSCAM-AS1* on oestrogen dependence in these cells, we witnessed an increased oestrogen dependence of T47D cells following *DSCAM-AS1* knockdown (Supplementary Fig. 7g).

To corroborate our *in vitro* findings in a tissue model, we obtained data previously generated performing ChIP-seq for ER in primary and metastatic breast tumour tissue[13]. These tumours were grouped into the following categories as previously described[13]: primary ER-negative ($n=2$), primary ER-positive, tamoxifen-responder ($n=8$), primary ER-positive, tamoxifen-non-responder ($n=7$), metastatic ER-positive ($n=3$). Strikingly, investigation of the *DSCAM-AS1* promoter revealed that ER preferentially binds to the *DSCAM-AS1* promoter in tumours with clinical aggression (ie, metastatic and tamoxifen non-responders; Fig. 5d), while a canonical ER target, *GREB1*, exhibits ER-biding to its promoter in nearly all ER-positive tumours, lacking preference for the more clinically aggressive

tumours. Altogether, these data suggest that the association between *DSCAM-AS1* expression with clinical aggressiveness in ER-positive breast cancer samples may be explained, in part, by the ability of *DSCAM-AS1* to facilitate oestrogen-independent oncogenicity, thus potentially promoting resistance to endocrine therapy with tamoxifen.

Discussion

Further investigation and study of the mechanisms through which ER-dominant breast cancers become aggressive and eventually evade traditional clinical therapies is of intense clinical interest. In this study, we identify a myriad of potentially ER-associated lncRNAs, and functionally and mechanistically characterize one of the most intriguing candidates. Nevertheless, further investigation of some of these other lncRNAs may also contribute to our understanding of ER biology and ER-driven oncogenesis. LncRNAs have been shown to function through multiple mechanisms, and the study of the interaction of *DSCAM-AS1* with hnRNPL is a promising step towards understanding the ways through which this molecule executes its oncogenic function. While we show that the binding of hnRNPL to *DSCAM-AS1* is responsible for at least some of its oncogenicity, a further understanding of how the interaction between hnRNPL and *DSCAM-AS1* is mediating this phenotype is necessary.

Novel mediators of tumour aggression, such as *DSCAM-AS1*, can provide insight into the mechanism of endocrine therapy resistance. This increased understanding may in turn lead to more effective strategies to overcome this resistance, which is one of the last, great clinical challenges in treatment of ER-positive breast cancer. In addition, there is little known regarding the role of noncoding RNAs in developing resistance to anti-oestrogen therapy, with a small number of studies implicating some of the more prominent, well characterized breast cancer lncRNAs[52,53]. *DSCAM-AS1* is just one of many potentially relevant ER-regulated lncRNAs in breast cancer, and further investigation of the other candidates is likely to yield a greater understanding of ER-mediated cancer biology. Ultimately, this study provides key insight into the role of lncRNAs in ER breast cancer biology, and is an important step in better understanding this common disease.

Methods

Cell lines and cell culture. All cell lines were obtained from the American Type Culture Collection (Manassas, VA). Cell lines were maintained using standard media and conditions. Specifically, T47D cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin and 5 mg ml^{-1} insulin. ZR75-1 cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin. HEK293 cells were maintained in DMEM plus 10% fetal bovine serum (FBS) plus 1% penicillin-streptomycin. MCF7 cells were cultured in Dulbecco's modified Eagle's medium plus GlutaMAX (DMEM, Invitrogen) containing 10% fetal bovine serum (Hyclone) and 1% penicillin-streptomycin. To establish the tamoxifen-resistant cell line, MCF7 cells were grown in IMEM phenol-red free medium with 10% Charcoal-stripped FBS in the presence of 1 uM (Z)-4-hydroxytamoxifen (Sigma) for 6 months. All cell lines were grown at 37 °C in a 5% CO_2 cell culture incubator, and were genotyped for identity at the University of Michigan Sequencing Core and tested routinely for *Mycoplasma* contamination.

Cell proliferation assay. Cells were seeded in a 48-well plate at 3×10^4 cells per well. Plates were added to Incucyte machine (Essen Bioscience) 16–20 h following seeding. Growth curves were constructed by imaging plates using the Incucyte system, where the growth curves are generated from confluence measurements acquired during continuous kinetic imaging. Four wells were measured per condition. For tamoxifen treatment, 16–20 h after seeding, the medium was changed to RPMI phenol-red free medium containing 10% charcoal-treated FBS in the presence of 1 uM tamoxifen or ethanol. Growth curves were obtained using Incucyte system as described above.

Cell viability assay. Cells were seeded in 96-well plates at 5,000 cells per well in a total volume of 100 µl media containing 10% FBS. Serially diluted tamoxifen in 100 µl of media was added to the cells 12 h after seeding. Medium containing tamoxifen was replenished every 2–3 days. Following 10 days of incubation, cell viability was assessed by WST assay (WST-8, Dojindo). All assays were performed in triplicate and repeated at least three times. The relative cell viability was expressed as a percentage of the control that was treated with vehicle solutions.

Soft agar colony formation assay. 10,000 cells were suspended in medium containing 0.3% agar, 10% FBS, and layered on medium containing 0.6% agar and 10% FBS in six-well plate. Colonies were stained for 18–24 h with iodonitrotetrozolium chloride (Sigma #18377) following 3 weeks of incubation. Colonies from three replicate wells were quantified.

Quantitative RT-PCR assay. The miRNeasy mini kit was utilized to isolate RNA from cell lysates. From 1 µg of isolated RNA, SuperScript III (Invitrogen) and Random Primers (Invitrogen) were used to generate cDNA according to the manufacturer's protocol. The ABI7900 HT Fast Real time system (Applied Biosystems) was utilized for quantitiative reverse transcriptase–PCR (qRT–PCR) reactions. Gene-specific primer were designed using the Primer3 software and were subsequently synthesized by IDT Technologies. A relative quantification method was used in analysing the qRT–PCR data and data were depicted as average fold change versus the control (as internal reference, GAPDH and actin were utilized). All primers used for qPCR are detailed in Supplementary Table 2. Three technical replicates were used in each assay, and all data shown was performed with at least three biological replicates.

Oestrogen and tamoxifen treatment. To evaluate the effect of oestrogen stimulation, cells were first hormone depleted via growth in phenol-red free medium containing 10% charcoal-treated FBS for 72 h and then treated with ethanol vehicle, 10 nM β-estradiol, or 10 nM β-estradiol plus 1 uM tamoxifen. After 10 h, RNA was isolated as described above and qPCR was performed as described above using Power SYBR Green Mastermix (Applied Biosystems).

Subcellular fractionation. Cellular fractionation was performed using a RiboTrap Kit (MBL International), according to manufacturer's instructions. RNA was isolated and qRT–PCR was performed as described above.

Knockdown and overexpression studies. Knockdown of *DSCAM-AS1* and *hnRNPL* in T47D and MCF7 cells was accomplished by small interfering RNA from Dharmacon. Transfections were performed with OptiMEM (Invitrogen) and RNAi Max (Invitrogen) per manufacturer instruction. Target sequences used for shRNA or small interfering RNA knockdown are listed in Supplementary Table 3. For stable knockdown of *DSCAM-AS1*, MCF7 and T47D cells were transfected with lentiviral constructs containing 2 different *DSCAM-AS1* shRNAs or no targeting shRNAs in the presence of polybrene (8 µg ml^{-1} Supplementary Table 3). After 48 h, transduced cells were grown in culture media containing 2 µg ml^{-1} puromycin.

For *DSCAM-AS1* overexpression, the predominant isoform (isoform 2, Supplementary Table 1) was cloned into the pLenti6.3 vector (Invitrogen) using PCR8 non-directional Gateway cloning (Invitrogen) as an initial cloning vector and shuttling was then done to pLenti6.3 using LR clonase II (Invitrogen) according to the manufacturer's instructions. As control, *LacZ* was also cloned into the same vector system. The primers for making *DSCAM-AS1* mutation and truncations are listed in Supplementary Table 2. Lentiviral particles were made and T47D and ZR75.1 cells were transduced as described above. Stable cell lines were generated by selection with 3 µg ml^{-1} blasticidin. Transient transfection of *DSCAM-AS1* and its derivative mutants was done in HEK293 cells was performed with Lipofectamine LTX (Invitrogen) following manufacturer's instruction. Cells were collected at 48 h post transfection.

***In vitro* RNA-binding assay.** The RNA-binding assay was performed according to the protocol of the RiboTrap Kit (MBL International). In brief, 5-bromo-UTP (BrU) was randomly incorporated into sense *DSCAM-AS1*, antisense *DSCAM-AS1*, and *LacZ* control via PCR-based transcription. The primers are shown in Supplementary Table 2. The the BrU labelled RNA transcripts were bound to beads conjugated with anti-BrdU antibodies. Then, the cytoplasmic or nuclear fractions from MCF7 or T47D cells were mixed for 2 h. Samples were washed four times with Wash Buffer II before elution. The samples were sent to the Michigan Center for Translational Pathology proteomic core facility for mass spectrometry.

Mass spectrometry. The samples were treated with SDS-PAGE loading buffer supplied with 10 mM DTT for 5 min at 85 °C. The proteins were alkylated by the addition of iodoacetamide to the final concentration of 15 mM. The samples were subjected to SDS–PAGE and the whole lanes were cut out and digested with trypsin in-gel for 2 h. The resulting peptides were extracted, dried and resuspended in 0.1% formic acid with 5% acetonitrile before loading onto a 2 cm EASY-column

(Thermo Scientific) coupled to an in-house made nano HPLC column (20 cm × 75 um) packed with LUNA C18 media (Phenomenex). Analysis was performed on a Velos Pro mass spectrometer (Thermo Scientific) operated in data-dependent mode using 120-min gradients in EASY-LC system (Proxeon) with 95% water, 5% acetonitrile (ACN), 0.1% formic acid (FA) (solvent A), and 95% ACN, 5% water, 0.1% FA (solvent B) at a flow rate of 220 nl min^{-1}. The acquisition cycle consisted of a survey MS scan in the normal mode followed by 12 data-dependent MS/MS scans acquired in the rapid mode. Charge state was not recorded. Dynamic exclusion was used with the following parameters: exclusion size 500, repeat count 1, repeat duration 10 s, exclusion time 45 s. Target value was set at 10^4 for tandem MS scan. The precursor isolation window was set at 2 *m/z*. The complete analysis comprised two independent biological replicates.

Mass spectrometry data analysis. The resulting spectrum files were transformed into MGF format by MSConvert software and interrogated by MASCOT 2.4 search engine using human UniProt database version 15 concatenated with reverse sequences for estimation of false discovery rate (FDR) and with a list of common contaminants (40,729 entries in total). The search parameters were as follows: full tryptic search, 2 allowed missed cleavages, peptide charges +2 and +3 only, MS tolerance 1 Da, MS/MS tolerance 0.5 Da. Permanent post-translational modification was: cysteine carbamidomethylation. Variable post-translational modifications were: protein N-terminal acetylation, Met oxidation and N-terminal Glutamine to pyro-Glutamate conversion. The remaining analysis was performed as previously described[54]. To summarize, the minimal ion score threshold was chosen such that a peptide FDR below 1% was achieved. The peptide FDR was calculated as: 2 × (decoy_hits)/(target + decoy hits). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[55] via the PRIDE partner repository with the data set identifier PXD002421 and 10.6019/PXD002421. Spectral counts for all detected proteins were assembled using an in-house written Python script. The adjustment of spectral counts was done as previously described[54].

RNA immunoprecipitation. RIP assays were performed using a Millipore EZ-Magna RIP RNA-Binding Protein Immunoprecipitation kit (Millipore, #17-700) according to the manufacturer's instructions. RIP-PCR was performed as qPCR, as described above, using total RNA as input controls. 1:150th of RIP RNA product was used per PCR reaction. Antibodies used for RIP are listed in Supplementary Table 4. All RIP assays were performed in biological duplicate.

Invasion assay. 3×10^5 cells were seeded in a 24-well corning FluoroBlok chamber pre-coated with Matrigel (BD Biosciences). Medium containing 10% FBS in the lower chamber served as chemoattractant. After 48 h, cells remaining on the lower side of the membrane were stained with calcein AM (C34852 invitrogen). The invasive cells adhering to the bottom surface of the filter were quantified under a fluorescent microscope (× 2).

Antibodies and immunoblot analyses. Western immunoblot assays were performed by running cell lysates on 4–12% SDS polyacrylamide gels (Novex) to separate proteins. Proteins were then transferred to a nitrocellulose membrane (Novex) via wet transfer at 30 V overnight. Blocking buffer incubation was then performed for 1 h (Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk). Indicated antibodies were then added to membrane and incubated at 4 °C overnight. Enhanced chemiluminescence (ECL Prime) was utilized to develop blots via the manufacturer's protocol. All the antibodies used in this study are described in Supplementary Table 4. Representative full blot images are shown in Supplementary Fig. 8.

Chromatin immunoprecipitation. HighCell# ChIP kit (Diagenode) was utilized to perform ChIP assays via the manufacturer's protocol. Briefly, MCF7 cells were grown in charcoal-stripped serum media (described above) for 72 h and then stimulated 10 nM estradiol for 12 h. Cells were then crosslinked using 1% formaldehyde for 10 min, and crosslinking was quenched for 5 min at room temperature using a 1/10 volume of 1.25 M glycine. Cells were then lysed and sonicated (Bioruptor, Diagenode), yielding an average chromatin fragment size of 300 bp. An equivalent amount of chromatin equivalent to 5×10^6 cells was used for the ChIP for all antibodies. DNA bound to immunoprecipitated product was isolated (IPure Kit, Diagenode) via overnight incubation with antibody at 4 °C. Samples were then washed, and crosslinked reversed.

ChIP-seq library construction and sequencing analysis. DNA was purified for library preparation using the IPure Kit (Diagenode). The ChIP-seq sample preparation for sequencing was performed according to the manufacturer's instructions (Illumina). ChIP-enriched DNA samples (1–10 ng) were converted into blunt-ended fragments using T4 DNA polymerase, *Escherichia coli* DNA polymerase I large fragment (Klenow polymerase) and T4 polynucleotide kinase (New England BioLabs (NEB)). A single adenine base was added to fragment ends by Klenow fragment (3′ to 5′ exo$^-$; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-modified DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (QIAGEN). Libraries were quantified with the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2000 Sequencer (100-nucleotide read length). ChIP-seq data were mapped to human genome version hg19 using BWA[56]. The MACS program[57] was used to generate coverage map files to visualize the raw signal on the UCSC genome browser[58]. Hpeak[59], a hidden Markov model (HMM)-based peak-calling software program designed for the identification of protein-interactive genomic regions, was used for ChIP-seq peak determination.

ChIP-seq peak promoter overlap. Overlap of ChIP-seq peaks with gene promoters was performed using the BEDTools 'coverage' tool. Intervals of ± 5–10 kilobases surrounding unique transcriptional starts were used to assess promoter overlap.

Coding potential scoring. Coding potential for all lncRNA transcripts was determined as described previously[4]. The alignment-free Coding Potential Assessment Tool (CPAT)[25] was used to determine coding probability for each transcript. CPAT determines the coding probability of transcript sequences using a logistic regression model built from ORF size, Fickett TESTCODE statistic, and hexamer usage bias.

Xenograft analysis. All experimental procedures were approved by the University of Michigan Committee for the Use and Care of Animals (UCUCA) and conform to all regulatory standards. A total of 5×10^6 cells of T47D control or T47D shM41 cells were suspended in 100 ul of PBS/Matrigel (1:1) were injected subcutaneously in 5-week-old pathogen-free female CB-17 severe combine immunodefiecient mice (CB-17 SCID) which simultaneously received a 60-day slow release pellet containing 0.18 mg of 17b-estradiol (Innovative Research of America). Tumours were measured weekly using a digital caliper. Growth in tumour volume was recorded using digital calipers and tumour volumes were estimated using the formula $(\pi/6)$ $(L \times W^2)$, where L = length of tumour and W = width. In addition, mouse livers were collected to determine spontaneous metastasis by measuring human Alu sequence. Briefly, genomic DNA from livers were prepared using Puregene DNA purification system (Qiagen), followed by quantification of human Alu sequence by human Alu specific Fluorogenic Taqman qPCR probes.

RNA-seq data processing. Sequence quality control was done using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Next, reads mapping to mitochondrial DNA, ribosomal RNA, poly-A, poly-C, Illumina sequencing adaptors, and the spiked-in phiX174 viral genome were filtered. Sequences were downloaded from the Illumina iGenomes server (2012, March 9). Mapping was performed using bowtie2 (2.0.2). Reads were mapped using TopHat2 (2.0.6 and 2.0.8) using default parameters. A human genome reference was constructed from UCSC version Feb 2009 (GRCh37/hg19) chromosomes 1–22, X, Y and mitochondrial DNA, and references from alternate haplotype alleles were omitted. Bowtie-build and bowtie2-build were used to build genome reference for Bowtie versions 0.12.8 and 2.0.2 were, respectively. The Ensembl version 69 transcriptome was used as a reference gene set. Using the --transcriptome-index option in TopHat version 2.0.6 (ref. 60), alignment index files were prepared from this reference for Bowtie versions 0.12.8 and 2.0.2.

RNA-seq transcript expression estimation. Cufflinks version 2.1.1 (ref. 61) was used with the following parameters to estimate transcript abundance from RNA-seq data: '--max-frag-multihits = 1', '--no-effective-length-correction', '--max-bundle-length 5000000', '--max-bundle-frags 20000000'. To convert FPKM abundance estimates (generated by Cufflinks) to approximate fragment count values we multiplied each FPKM by the transcript length (in kilobases) and by the 'Map Mass' value (divided by 1.0E^6) found in the Cufflinks log files.

Breast cancer tissue expression heatmap generation. The 'gplots' R-package was used to generate heatmaps using the *heatmap.2* function. For the cancer versus normal heatmap, expression was normalized as log2 of the fold change over the median of the normal samples for each transcript. For the ER-positive versus ER-negative heatmap, expression was normalized to the median of the ER-negative samples. Unsupervised heirarchical clustering was performed with the *hclust* function, using Pearson correlation as the clustering distance, using the 'ward' agglomeration method.

RNA-seq differential expression testing. Differential expression testing was performed using the SSEA tool described previously[7]. Briefly, following count data normalization, SSEA performs the weighted KS-test procedure described in GSEA[42]. The resulting enrichment score statistic describes the enrichment of the sample set among all samples being tested. To test for significance, SSEA enrichment tests are performed following random shuffling of the sample labels.

These shuffled enrichment tests are used to derive a set of null enrichment scores (1,000 null enrichment scores computed). The nominal *P* value reported is the relative rank of the observed enrichment score within the null enrichment scores. Multiple hypothesis testing is performed by comparing the enrichment score of the test to the null normalized enrichment score distributions for all transcripts in a sample set. This null normalized enrichment score distribution is used to compute FDR *Q* values in the same manner used by GSEA[42].

Associations with oncomine clinical signatures. We identified the top 150 positively and negatively correlated genes (Spearman's correlation) to *DSCAM-AS1* among the ER-positive breast cancer samples. These gene lists were imported into Oncomine[27] as custom concepts. We then nominated significantly associated breast cancer concepts with odds ratio >4.0 for negatively associated concepts and >6.0 for positively associated concepts and *P* value $<1\times10^{-6}$. Nodes and edges of these associations were exported and a concept association network was generated using Cytoscape version 3.2.1. Node positions were computed using the Force-Directed Layout algorithm in Cytoscape using the odds ratio as the edge weight. Node positions were subtly altered manually to enable better visualization of node labels.

Association of correlation signatures with oncomine concepts. Correlation analysis described above was performed for *DSCAM-AS1*, *EZH2*, *HOTAIR*, *MALAT1*, and *NEAT1*. For each gene, we created a signature of the top 150 most positively and top 150 most negatively correlated genes. We performed a Fisher's exact test of overlap for each of the above gene signatures with Oncomine clinical signatures for cancer versus normal, clinical recurrence, clinical survival, metastasis, and high clinical grade. The following studies were utilized: Curtis Breast[26], Ma Breas[62], TCGA Breast[28], Zhao Breast[29], Bittner Breast[63], Desmedt Breast[30], Ivshina Breast[31], Loi Breast[32], Lu Breast[33], Perou Breast[22], Schmidt Breast[34], Sorlie Breast[23], vantVeer Breast[64], Wang Breast[36], Boersma Breast[37], Kao Breast[38], Symmans Breast[39] and vandeVijver Breast[40]. For each Oncomine concept, overlap was tested for the top 1, 5 and 10% of genes up- and downregulated, and the gene signature with the greatest odds ratio was selected for each study. Signature comparisons were performed using a one-sided Fisher's exact test.

Survival analysis with TCGA breast data. Association of DSCAM-AS1 levels on clinical outcomes was assessed using the TCGA breast cohort. Survival data was obtained from the TCGA data portal. ER-positive samples were used for survival analysis as indicated by the TCGA clinical metadata via IHC status. Samples with DSCAM-AS1 expression >10 FPKM were grouped into the 'DSCAM-AS1 high' category and samples with expression <1 FPKM were grouped into the 'DSCAM-AS1 low' category. Kaplan-Meier analysis was performed, and log-rank test was performed to assess statistical significance.

Tissue expression level percentile metric. To generate a metric to summarize the expression of each lncRNA in breast cancer tissues, we identified the expression level of the 95th percentile sample among all breast RNA-seq samples including cancers tissue, normal tissue, and cell lines.

RNA-sequencing library preparation. Total RNA was obtained from cancer cell lines, and RNA quality was determined using an Agilent Bioanalyzer. Poly-A transcriptome libraries from the mRNA fractions were generated following the Illumina RNA-seq protocol. Each sample was sequenced in a single lane with the Illumina HiSeq 2000 (100-nucleotide read length) as previously described[3,65]. The dUTP method of second-strand marking was used for strand-specific library preparation as described previously[66].

Gene set enrichment analysis. Expression levels of *DSCAM-AS1* were correlated (Spearman) to the expression of all protein-coding genes across all ER-positive breast cancers. The protein-coding genes were then ranked by the Spearman Rho value, and used in a weighted, preranked GSEA analysis against MSigDB gene sets V5.0 (ref. 67).

***In silico* binding prediction.** To obtain potential HNRNPL binding sites on DSCAM-AS1, we utilized GraphProt[68] to learn a predictive model from genome-wide HNRNPL binding sites identified by iCLIP-seq[48]. For training data generation, we extracted the genomic binding positions (GSE37560) with BED table scores $>=10$, followed by an extension of ±20 nt resulting in 41 nt long binding sites. After mapping the sites to annotated RefSeq genes obtained from UCSC, an equally-sized set of negative sites was selected such that the sites were on the same RefSeq genes and did not overlap with any of the identified positive sites from the initial table. The GraphProt sequence model trained on these data was then used to identify high-scoring sites in the *DSCAM-AS1* sequence (NCBI GenBank NR_038899.1). The highest-scoring site centred at RNA position 923 contains a CA-repeat motif known for its affinity towards HNRNPL and was thus used for subsequent analysis.

Rapid amplification of cDNA ends (RACE). 5′ and 3′ RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen) according to the manufacturer's protocols. RACE PCR products obtained using Platinum Taq high-fidelity polymerase (Invitrogen), were resolved on a 1.5% agarose gel. Individual bands were gel purified using a Gel Extraction kit (Qiagen), and cloned into PCR4 TOPO vector, and sequenced using M13 primers.

Single-molecule fluorescence *in situ* hybridization. Single-molecule fluorescence *in situ* hybridization was performed as described[69], with some minor modifications. Cells were grown on 8-well chambered coverglasses, formaldehyde fixed and permeablized overnight at 4 °C using 70% ethanol. Cells were rehydrated in a solution containing 10% formamide and $2\times$ SSC for 5 min and then treated with 10 nM fluorescence *in situ* hybridization probes for 16 h in $2\times$ SSC containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNAse-free BSA, 1 μg μl^{-1} *E. coli* transfer RNA and 10% formamide at 37 °C. After hybridization, cells were washed twice for 30 min at 37 °C using a wash buffer (10% formamide in $2\times$ SSC). Cells were then mounted in solution containing 10 mM Tris/HCl pH 7.5, $2\times$ SSC, 2 mM trolox, 50 μM protocatechiuc acid and 50 nM protocatechuate dehydrogenase. fluorescence *in situ* hybridization samples were imaged in three dimensions using HILO illumination as described[70]. Images were processed using custom-written macros in ImageJ. Analysis routines comprises 3 major steps: background subtraction, Laplacian of Gaussian (LoG) filtering and thresholding. Spots with intensity above set threshold are represented in images. Probes were designed to target all isoforms of the DSCAM-AS1 transcript. Probe sequences targeting DSCAM-AS1 (21 probes per transcript) are as follows: 5′-cctatccctttctctaagaa-3′, 5′-acttctgcaaaaacgtgctg-3′, 5′-ggttccactccatt ttaatt-3′, 5′-ctatagcgtcttatcagctg-3′, 5′-catgtgtccggatatcattt-3′, 5′-tcagtgagtggataact ggt-3′, 5′-aattctagtggaggcaccta-3′, 5′-ctaagtagcttcatctttcc-3′, 5′-caactgcgtgtttccta gtc-3′, 5′-agcattctctgttttaacca-3′, 5′-ttagcaactgccttgctctg-3′, 5′-gctgtccagttttagta aca-3′, 5′-cgttgtgagcctgagagatc-3′, 5′-agaacttccctagaggagtg-3′, 5′-atggggagtgagaccaa aca-3′, 5′-tggaggagggacagagaagg, 5′-tgtgggtgattggtactttt-3′, 5′-atggatgagtatgtcat gcc-3′, 5′-tattgccatggttagcatga-3′, 5′-aatgcatgcttgatggagct-3′.

Data availability. Sequence data that support the findings of this study have been deposited in the Short Read Archive with the accession code SRP078392. Tissue ChIP-seq data referenced in this study are available in the Gene Expression Omnibus with the accession code GSE32222. All remaining data are contained within the Article and Supplementary Information files or available from the author on request.

References

1. Prensner, J. R. & Chinnaiyan, A. M. The Emergence of lncRNAs in Cancer Biology. *Cancer Discov.* **1,** 391–407 (2011).
2. Ulitsky, I. & Bartel, D. P. LincRNAs: Genomics, evolution, and mechanisms. *Cell* **154,** 26–46 (2013).
3. Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* **29,** 742–749 (2011).
4. Prensner, J. R. *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45,** 1392–1398 (2013).
5. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142,** 409–419 (2010).
6. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–1076 (2010).
7. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47,** 199–208 (2015).
8. Cancer Genome Atlas Research, N. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2015).
9. Sun, M., Gadad, S. S., Kim, D. S. & Kraus, W. L. Discovery, annotation, and functional analysis of long noncoding RNAs controlling cell-cycle gene expression and proliferation in breast cancer cells. *Mol. Cell* **59,** 698–711 (2015).
10. Sotiriou, C. *et al.* Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl Acad. Sci. USA* **100,** 10393–10398 (2003).
11. Haque, R. *et al.* Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol. Biomarkers Prev.* **21,** 1848–1855 (2012).
12. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38,** 1289–1297 (2006).
13. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481,** 389–393 (2012).
14. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45,** 1446–1451 (2013).
15. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* **45,** 1439–1445 (2013).

16. Osborne, C. K. & Schiff, R. Mechanisms of endocrine resistance in breast cancer. *Annu. Rev. Med.* **62,** 233–247 (2011).
17. Qiu, J.-J. *et al.* Expression and clinical significance of estrogen-regulated long non-coding RNAs in estrogen receptor α-positive ovarian cancer progression. *Oncol. Rep.* **31,** 1613–1622 (2014).
18. Chakravarty, D. *et al.* The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **5,** 5383 (2014).
19. Bhan, A. *et al.* Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. *J. Mol. Biol.* **425,** 3707–3722 (2013).
20. Miano, V. *et al.* Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. *Oncotarget* **7,** 3201–3216 (2016).
21. Hah, N. *et al.* A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145,** 622–634 (2011).
22. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
23. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98,** 10869–10874 (2001).
24. Robinson, D. R. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat. Med.* **17,** 1646–1651 (2011).
25. Wang, L. *et al.* CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41,** 1–7 (2013).
26. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486,** 346–352 (2012).
27. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform1. *Neoplasia* **6,** 1–6 (2004).
28. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
29. Zhao, H. *et al.* Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell* **15,** 2523–2536 (2004).
30. Desmedt, C. *et al.* Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin. Cancer Res.* **14,** 5158–5165 (2008).
31. Ivshina, A. V. *et al.* Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66,** 10292–10301 (2006).
32. Loi, S. *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.* **25,** 1239–1246 (2007).
33. Lu, X. *et al.* Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res. Treat.* **108,** 191–201 (2008).
34. Schmidt, M. *et al.* The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68,** 5405–5413 (2008).
35. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
36. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365,** 671–679 (2005).
37. Boersma, B. J. *et al.* A stromal gene signature associated with inflammatory breast cancer. *Int. J. Cancer* **122,** 1324–1332 (2008).
38. Kao, K.-J., Chang, K.-M., Hsu, H.-C. & Huang, A. T. Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* **11,** 143 (2011).
39. Symmans, W. F. *et al.* Genomic index of sensitivity to endocrine therapy for breast cancer. *J. Clin. Oncol.* **28,** 4111–4119 (2010).
40. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
41. Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl Acad. Sci. USA* **100,** 11606–11611 (2003).
42. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
43. Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. & Jang, S. K. Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.* **298,** 395–405 (2000).
44. Liu, X. & Mertz, J. E. HnRNP L binds a cis-acting RNA sequence element that enables intron- independent gene expression. *Genes Dev.* **9,** 1766–1780 (1995).
45. Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat. Struct. Biol.* **10,** 33–37 (2003).
46. Rothrock, C. R., House, A. E. & Lynch, K. W. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J.* **24,** 2792–2802 (2005).
47. Hung, L.-H. *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14,** 284–296 (2008).
48. Rossbach, O. *et al.* Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L. *RNA Biol.* **11,** 146–155 (2014).
49. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
50. Mouridsen, H. *et al.* Letrozole therapy alone or in sequence with tamoxifen in women with breast cancer. *N. Engl. J. Med.* **361,** 766–776 (2009).
51. Winer, E. P. *et al.* American Society of Clinical Oncology technology assessment on the use of aromatase inhibitors as adjuvant therapy for postmenopausal women with hormone receptor-positive breast cancer: status report 2004. *J. Clin. Oncol.* **23,** 619–629 (2005).
52. Xue, X. *et al.* LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* **35,** 2746–2755 (2015).
53. Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L. C. J. & Van Agthoven, T. Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell. Physiol.* **226,** 1741–1749 (2011).
54. Poliakov, A. *et al.* Large-scale label-free quantitative proteomics of the pea aphid-Buchnera symbiosis. *Mol. Cell. Proteomics* **10,** M110.007039 (2011).
55. Vizcaíno, J., Deutsch, E. & Wang, R. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32,** 223–226 (2014).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
57. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
58. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42,** 764–770 (2014).
59. Qin, Z. S. *et al.* HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. *BMC Bioinformatics* **11,** 369 (2010).
60. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
61. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–578 (2012).
62. Ma, X.-J. *et al.* Gene expression profiles of human breast cancer progression. *Proc. Natl Acad. Sci. USA* **100,** 5974–5979 (2003).
63. McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
64. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
65. Maher, C. A. *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl Acad. Sci. USA* **106,** 12353–12358 (2009).
66. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7,** 709–715 (2010).
67. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).
68. Maticzka, D., Lange, S. J., Costa, F. & Backofen, R. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol.* **15,** R17 (2014).
69. Raj, A., van den Bogaard, P., Rifkin, S. a., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5,** 877–879 (2008).
70. Pitchiaya, S., Androsavich, J. R. & Walter, N. G. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Rep.* **13,** 709–715 (2012).

Acknowledgements

We thank K.W.-R. for help with the mouse Xenograft work, the University of Michigan DNA Sequencing Core for Sanger sequencing and K. Giles for critically reading the manuscript and for the submission of documents. This work was supported in part by US National Institutes of Health Prostate Specialized Program of Research Excellence grant P50CA186786, Early Detection Research Network grant UO1 CA111275, US National Institutes of Health grants R01 CA132874 and RO1 CA154365 (A.M.C.), and US Department of Defense grant PC100171 (A.M.C.). A.M.C. is supported by the Prostate Cancer Foundation and the Howard Hughes Medical Institute. A.M.C. is an American Cancer Society Research Professor and a Taubman Scholar of the University of Michigan. R.M. was supported by a Prostate Cancer Foundation Young Investigator Award and by US Department of Defense Post-Doctoral Fellowship W81XWH-13-1-0284. Y.S.N. is supported by a University of Michigan Cellular and Molecular Biology National Research Service Award Institutional Predoctoral Training Grant.

Author contributions

Y.S.N., S.H., A.M.C. and F.Y.F. conceived the study and analyses. Y.S.N. performed all bioinformatics analyses and guided study design with the assistance of M.K.I. and S.G.Z. S.H. performed cellular and molecular biology experiments with the assistance T.M., R.M., C.Z., J.R.P., Y.H., S.R., S.K., R.F.S., L.X. and U.S., K.W.-R. performed mouse xenograft work. X.C. generated RNA-seq libraries and performed the sequencing. A.P. performed the mass spectrometry. M.U., A.G., C.C., R.B. and C.S.S. performed the *in silico* binding predictions. J.M.R. produced the tamoxifen resistance cell line. D.F.H. and L.J.P. provided guidance and analysis for the clinical work. Y.S.N., A.M.C.

and F.Y.F. wrote the manuscript. All authors discussed results and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: Oncomine is supported by ThermoFisher, Inc. (Previously Life Technologies and Compendia Biosciences). A.M.C. was a co-founder of Compendia Biosciences and served on the scientific advisory board of Life Technologies before it was acquired. D.F.H. receives research funding from Astra-Zeneca, Eli Lilly, Janssen, Pfizer and Puma. He has stock options in OncImmune and in InBiomotion, and he is named as a primary inventor on two patents issued to the University of Michigan, one of which is licensed to Janssen and for which he receives royalties. These patents are not pertinent to the work presented here. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Niknafs, Y. S. *et al.* The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat. Commun.* 7:12791 doi: 10.1038/ncomms12791 (2016).

Lanbo Xiao, Ph.D.
Research Fellow

1400 E Medical Center Drive
Room No. 5331, Comprehensive Cancer Center
University of Michigan, Ann Arbor, MI-48105
Cell Phone: (734) 2223 7146; E-mail: lanbox@med.umich.edu

Education and Training

September 2004 – June 2008	BS in Biotechnology, Hunan Agriculture University, Hunan, China
September 2008 – December 2013	Ph.D. in Physiopathology, Central South University, Hunan, China

Research Experience

September 2008 – December 2013	Graduate Student (PhD) at the Central South University
July 2014 – Present	Research Fellow at the University of Michigan

Research Interest

• Discovery and characterization of novel long non-coding RNA associated with prostate cancer
• To evaluate lncRNAs as prognostic and diagnostic biomarkers.
• Identification and therapeutic targeting of novel androgen receptor co-activators in lethal castration resistant prostate cancer.

Grants/Research Support

W81XWH-16-1-0195
Department of Defense Post-Doctoral Fellowship June/16-Present
Investigator
$61,477/yr

Title: Biological characterization and clinical utilization of metastatic prostate cancer-associated lincRNA *SChLAP1*

Peer-Reviewed Publications

1. Shukla S*, Zhang X*, Niknafs YS*, **Xiao L***, Mehra R, Cieślik M, Ross A, Schaeffer E, Malik B, Guo S, Freier SM, Bui HH, Siddiqui J, Jing X, Cao X, Dhanasekaran SM, Feng FY, Chinnaiyan AM, Malik R. Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. Neoplasia. 2016 Aug;18(8):489-99.
2. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y, Prensner JR, Poliakov A, Singhal U, **Xiao L**, Kregel S, Siebenaler RF, Zhao SG, Uhl M, Gawronski A, Hayes DF, Pierce LJ, Cao X, Collins C, Backofen R, Sahinalp CS, Rae JM, Chinnaiyan AM, Feng FY. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. Nat Commun. 2016 Sep 26;7:12791.
3. Yang L, Liu L, Xu Z, Liao W, Feng D, Dong X, Xu S, **Xiao L**, Lu J, Luo X, Tang M, Bode AM, Dong Z, Sun L, Cao Y. EBV-LMP1 targeted DNAzyme enhances radiosensitivity by inhibiting tumor angiogenesis via the JNKs/HIF-1 pathway in nasopharyngeal carcinoma. Oncotarget. 2015 Mar 20;6(8):5804-17.
4. Yu X, Deng Q, Li W, **Xiao L**, Luo X, Liu X, Yang L, Peng S, Ding Z, Feng T, Zhou J, Fan J, Bode AM, Dong Z, Liu J, Cao Y. Neoalbaconol induces cell death through necroptosis by regulating RIPK-dependent autocrine TNFα and ROS production. Oncotarget. 2015 Feb 10;6(4):1995-2008.
5. **Xiao L**, Hu ZY, Dong X, Tan Z, Li W, Tang M, Chen L, Yang L, Tao Y, Jiang Y, Li J, Yi B, Li B, Fan S, You S, Deng X, Hu F, Feng L, Bode AM, Dong Z, Sun LQ, Cao Y. Epstein-Barr Virus-Mediated Glycolysis Enhances Resistance to Radiation Therapy in Nasopharyngeal Carcinoma. Oncogene. 2014 Sep 11;33(37):4568-78.
6. Li W, Peng C, Lee MH, Lim D, Zhu F, Fu Y, Yang G, Sheng Y, **Xiao L**, Dong X, Ma W, Bode AM,Cao Y, Dong Z. TRAF4 is a critical molecule for Akt activation in lung cancer. Cancer Res. 2013 Dec 1;73(23):6938-50.
7. Deng Q, Yu X, **Xiao L**, Hu Z, Luo X, Tao Y, Yang L, Liu X, Chen H, Ding Z, Feng T, Tang Y, Weng X,Gao J, Yi W, Bode AM, Dong Z, Liu J, Cao Y. Neoalbaconol Induces Energy Depletion and Cell Death in Cancer Cells by Targeting PDK1-PI3-K/Akt Signaling Pathway. Cell Death Dis. 2013 Sep 19;4:e804.

8. Li L*, Li W*, **Xiao L**, Xu J, Chen X, Tang M, Dong Z, Tao Q, Cao Y. Viral oncoprotein LMP1 disrupts p53-induced cell cycle arrest and apoptosis through modulating K63-linked ubiquitination of p53. Cell Cycle. 2012 Jun 15;11(12):2327-36.
9. Shi Y, Tao Y, Jiang Y, Xu Y, Yan B, Chen X, **Xiao L**, Cao Y. Nuclear epidermal growth factor receptor interacts with transcriptional intermediary factor 2 to activate cyclin D1 gene expression triggered by the oncoprotein latent membrane protein 1. Carcinogenesis. 2012 Aug;33(8):1468-78.
10. Guo L, Tang M, Yang L, **Xiao L**, Bode AM, Li L, Dong Z, Cao Y. Epstein-Barr virus oncoprotein LMP1 mediates survivin upregulation by p53 contributing to G1/S cell cycle progression in nasopharyngeal carcinoma. Int J Mol Med. 2012 Apr;29(4):574-80.
11. Ma X*, Yang L*, **Xiao L**, Tang M, Liu L, Li Z, Deng M, Sun L, Cao Y. Down-regulation of EBV-LMP1 radio-sensitizes nasal pharyngeal carcinoma cells via NF-κB regulated ATM expression. PLoS One. 2011;6(11):e24647.